Editors: Richard Sharp, Rebecca Chaplin-Kramer, Spencer Wood, Anne Guerry, Heather Tallis, Taylor Ricketts.

Contributing Authors: Erik Nelson, Driss Ennaanay, Stacie Wolny, Nasser Olwero, Kari Vigerstol, Derric Pennington, Guillermo Mendoza, Juliann Aukema, John Foster, Jessica Forrest, Dick Cameron, Katie Arkema, Eric Lonsdorf, Christina Kennedy, Gregory Verutes, Chong-Ki Kim, Gregory Guannel, Michael Papenfus, Jodie Toft, Matthew Marsik, Joey Bernhardt, Robert Griffin, Kathryn Glowinski, Nicholas Chaumont, Adam Perelman, Martin Lacayo, Lisa Mandle, Perrine Hamel, Adrian L. Vogl, Lauren Rogers, Will Bierbower, Douglas Denu.

Citation: Sharp, R., Tallis, H.T., Ricketts, T., Guerry, A.D., Wood, S.A., Chaplin-Kramer, R., Nelson, E., Ennaanay, D., Wolny, S., Olwero, N., Vigerstol, K., Pennington, D., Mendoza, G., Aukema, J., Foster, J., Forrest, J., Cameron, D., Arkema, K., Lonsdorf, E., Kennedy, C., Verutes, G., Kim, C.K., Guannel, G., Papenfus, M., Toft, J., Marsik, M., Bernhardt, J., Griffin, R., Glowinski, K., Chaumont, N., Perelman, A., Lacayo, M. Mandle, L., Hamel, P., Vogl, A.L., Rogers, L., and Bierbower, W. 2015. InVEST +VERSION+ User's Guide. The Natural Capital Project, Stanford University, University of Minnesota, The Nature Conservancy, and World Wildlife Fund.

CONTENTS

Part I

Introduction

CHAPTER

ONE

DATA REQUIREMENTS AND OUTPUTS SUMMARY TABLE

InVEST Data and Model Inventory				
	Step	**Data requirements**	**Process**	**Outputs**
Biodiversity: Habitat Quality and Rarity (Tier 0)				
Required	**Supply**	Current Land use/land cover	Calculate habitat quality and degradation based on threat intensity and sensitivity	Habitat degradation index; Habitat quality index
		Threat impact distance		
		Relative threat impact weights		
		Form of threat decay function		
		Threat maps		
		Habitat suitability (optional: by species group)		
		Habitat sensitivity to threats		
		Half saturation constant		
Optional	**Supply**	Protected status		
		Baseline land use/land cover	Calculates rarity of current and/or future habitat types relative to baseline; calculates quality and degradation of baseline based on threat intensity and sensitivity	Relative habitat rarity index for current and/or future land use/land cover; Degradation and quality for baseline
		Future land use/land cover	Calculates quality and degradation of future scenario based on threat intensity and sensitivity; optionally calculates habitat rarity relative to baseline	Habitat degradation, quality and optionally rarity for future scenario
Carbon Storage and Sequestration				
Required	**Service**	Land use/land cover	Looks up carbon stock(s) per pixel	Total carbon stock (Mg/pixel)
		Carbon in aboveground biomass		
		Carbon in belowground biomass		
		Carbon in dead organic matter		
		Carbon in soil		
Optional	**Service**	Carbon removed via timber harvest	Calculates carbon stored in harvested wood products per pixel	Total carbon stock, including that in HWP (Mg/pixel)
		First year of timber harvest		
		Harvest frequency		
		Half life of harvested wood products		
		Carbon density in harvested wood		
		Biomass conversion expansion factor		
		Future land use/land cover	Calculates difference between carbon stocks	Carbon sequestration rates (Mg/pixel/yr)
Optional	**Value**	Value of sequestered carbon	Calculates value of carbon	Value of sequestered carbon (currency/pixel/yr)
		Discount rate		
		Timespan		
		Annual rate of change in price of carbon		
Hydropower Production (Tier 1)				
Required	**Supply**	Land use/land cover	Calculates pixel level yield as difference between precipitation and actual evapotranspiration	Mean annualyield (mm/watershed/yr, mm/pixel/yr)
		Mean annual precipitation (mm)		
		Mean annual reference evapotranspiration (mm)		
		Plant available water content (fraction)		
		Evapotranspiration coefficient		
		Root depth (mm)		
		Effective soil depth (mm)		
		Seasonality factor		
Required	**Service**	Consumptive use by LULC	Subtracts water consumed forby different land use and cover	Mean annual water yield available for hydropower production (mm/watershed/yr.
		Subwatershed and Watershed shapefiles		
		Calibration coefficient		
Optional	**Value**	Turbine efficiency (0.7-0.95)	Estimates power for a given volume of water	Energy production (KWH/watershed/yr, KWH/pixel/yr)
		Inflow volume for hydropower (fraction)		
		Hydraulic head (m)		
		Operation cost (currency)	Calculates net present value of energy	Net present value (currency/watershed/yr
		Hydropower price (currency)		

Water Purification: Nutrient Retention (Tier 1)

Required	**Supply**	Land use/land cover DEM Water yield (output from Hydropower model; refer to Hydropower model for input data requirements) Export coefficient in kg/ha/yr (for nutrient(s) of interest) Nutrient filtration efficiency (%)	Calculates nutrient export and retention	Nutrient export (kg/watershed/yr, kg/pixel/yr) Nutrient retention (kg/watershed/yr, kg/pixel/yr)
Required	**Service**	Allowed level of nutrient pollution Subwatershed and watersheds shapfiles	Subtracts retention equal to amount of allowed pollution	water purification through ecosystem nutrient retention (kg/watershed/yr, kg/pixel/yr)
Optional	**Value**	Mean annual nutrient removal costs Lifespan (years) Discount rate (%)	Calculates present value of costs	Avoided treatment costs (currency/watershed/yr,c basin/yr)

Sediment Retenion Model: Avoided Dredging and Water Quality Regulation (Tier 1)

Required	**Supply**	Land use/land cover Rainfall erosivity Soil erodibility Crop factor Management factor DEM Sediment retention efficiency for each LULC Slope threshold (%) Flow accumulation threshold	Calculates generated and retained sediment at pixel scale using USLE and routing	Mean annual erosion (tons/watershed/yr, tons/pixel/yr) Mean annual sediment retention (tons/watershed/yr, tons/pixel/yr)
Required	**Reservoir Service**	Reservoir dead volume (reservoir points of interest) Subwatershed and Watershed shapfiles	Subtracts sediment loads in reservior dead volume	Mean annual generated and retained sediment loads (tons/watershed/yr,
Required	**Treatment Plant Service**	Allowed sediments load in rivers (TMDL, etc.)	Subtracts sediment loads equal to allowed load	Annual average sediment retention of value to water treatment plants
Optional	**Avoided Dredge Value**	Mean annual dredging cost (Currency) Lifespan (years) Discount rate (%)	Calculates present value of dredging costs	Avoided dredge costs (currency/watershed/yr, currency/pixel/yr)
	Avoided Treatment Value	Mean annual sediment removal cost (Currency) Lifespan (years) Discount rate (%)	Calculates present value of treatment costs	Avoided treatment costs (currency/watershed/yr, currency/pixel/yr)

Managed Timber Production (Tier 1)

Required	**Service**	Location of timber parcels Area per timber parcel Proportion of timber harvested per parcel per period Wood biomass harvested per parcel per period Harvest period per parcel Harvested wood mass:volume conversion factor	Calculates amount of timber harvested	Harvested timber volume (m3/parcel/yr) Harvested timber biomass (Mg/parcel/yr)
Optional	**Value**	Market price of timber Annual average plantation maintenance costs Annual average harvest costs Timeframe into future harvests will be valued Discount rate	Calculates net present value of timber harvested	Net present value of timber (currency/parcel/yr)

Crop Pollination (Tier 0)

Required	**Supply**	Land Use/Land Cover Nesting Habitat Preference Relative Index of seasonal pollinator activity Relative availability of nesting habitat types Relative abundance of flowers per LULC Average foraging distance	Calculates relative abundance of pollinators	Index of pollinator abundance (relative abundance/pixel, relative abundance/watershed)
Required	**Service**	Relative abundance index (supply from above)	Calculates relatve abundance of pollinators visiting each farm	Index of relative pollinator abundance on farms (relative abundance/farm)
Optional	**Value**	**Crop half saturation constant**	Calculates relative additional value of pollination	Index of crop yield value from pollination (relative value/pixel)

Food from Fisheries

Required	**Supply**	spatial structure life history traits: age/stage-specific survival, fecundity, age structure productivity (R/S) fishing mortality rate (age/stage specific)	estimates adult abundance available for harvest or escapement; trend (lambda) of returns	Number of total returns (escapement + catch) or escaped spawners per year or trend in returns or escapement
Required	**Service**	harvest management strategy: 1. sector-specific catch or harvest rate; or 2. target escapement and sector-specific allocation	estimates number of landed fish from each population	number of fish landed per year by sector (commercial and subsistence)
Optional	**Value**	Annual average sediment removal cost market price operating costs	Calculates present value of fish landed	net present value of fish landed by sector

Food from Aquaculture

Required	**Service**	farm operations (number of fish, feed, target harvest weight, weight at outplanting, date of outplanting, fallowing practices) farm locations temperature	estimates biomass of fish produced per farm	Biomass of fish produced per farm
Optional	**Value**	operating costs market price revenues	Calculates present value of fish produced per farm	net present value of fish produced per farm

Protection from coastal erosion

Required	**Supply**	wind field wave field bathymetry tides shoreline type/backshore characteristics benthic biogenic habitats topography (optional)	calculates attenuated wave height; calculates total water level from run-up (via wave height and wave set-up) and/or via storm surge; calculates cross-shore erosion	area of shoreline lost per storm event
Required	**Service**	Land use/land cover map (location of properties or beaches eroded) location and type of infrastructure placed in nearshore region	calculates avoided loss of beach or shoreline w&w/o biogenic habitat	avoided loss of property or infrastructure per event (private property); avoided loss of beach per event (private property)
Optional	**Value**	value of property or beaches eroded Beach carrying capacity beach nourishment costs value of infrastructure eroded shoreline hardening costs	Calculates present value of damage per event, of a beach visitor, beach nourishment or shoreline hardening.	value of avoided property or infrastructure damage per event; value of avoided dune nourishment or shoreline protection; value of avoided tourist revenue lost

Protection from coastal inundation

Required	**Supply**	wind field	calculates attenuated wave height; calculates total water level from run-up (via wave height and wave set-up) and/or via storm surge; calculates stability of sand dunes	area of property or infrastructure flooded per event
		wave field		
		bathymetry		
		tides		
		shoreline type/backshore characteristics		
		benthic biogenic habitats		
		topography		
Required	**Service**	Land use/land cover map (location of properties or beaches eroded)	calculates avoided area flooded w&w/o biogenic habitat	avoided property or infrastructure damage per event
		location and type of infrastructure placed in nearshore region		
Optional	**Value**	value of property inundated	Calculates present value of flooded area or infrastructure, costs of dune nourishment or added shoreline protection	value of avoided property or infrastructure damage per event; value of avoided dune nourishment or shoreline protection
		dune nourishment costs		
		value of infrastructure inundated		
		man-made shoreline protection construction costs		

Wave energy generation

Required	**Supply**	wave height	calculates wave power resource from wave data	wave power resource at each location
		wave period		
		bathymetry		
		tides		
Required	**Service**	device attributes (conversion efficiency)	calculates captured wave energy	captured wave energy per array (MWh)
		array design		
		array location		
Optional	**Value**	capital costs (e.g., device, cables, etc.)	Calculates present value of electricity captured per array	net present value of electricity captured from waves per array
		operating costs		
		revenue		
		life span of array facility		

Aesthetic value from viewsheds

Required	**Supply**	attributes of marine environment (location of natural desired features & development/infrastructure)	calculates points from which natral/desired or infrastructure can be observed	
		attributes of shoreline environment (location of natural desired features & development/infrastructure)		
		bathymetry		
		topography		
Required	**Service**	access points	calculates points from which infrastructure can be observed	number of natural (non-infrastructure or development) views per location
		location of public parks		
		location of private property		

Recreation Value				
Required	**Supply**	location of natural desired features for recreation (e.g., whale sightings, mammal haul outs, kelp for SCUBA, beaches, etc.)	maps locations of recreation activities	
		location and quality of environmental conditions affecting recreation value (e.g., wave energy for beach enjoyment or wildlife viewing)		
Required	**Service**	location of infrastructure in support of recreation activities (e.g., campgrounds, boat launches, etc.)	calculates index of recreation importance	index of recreation importance by activity and weighted overall index
		distance between access points and activities		
		visitation rates for each location, activity		
Optional	**Value**	visitation for each activity	Calculates present value of electricity captured per array	net present value of electricity captured from waves per array
		travel costs		
		revenue from activities		

CHAPTER

TWO

WHY WE NEED TOOLS TO MAP AND VALUE ECOSYSTEM SERVICES

2.1 Introduction

Ecosystems, if properly managed, yield a flow of services that are vital to humanity, including the production of goods (e.g., food), life support processes (e.g., water purification), and life fulfilling conditions (e.g., beauty, recreation opportunities), as well as the conservation of options (e.g., genetic diversity for future use). Despite its importance, this natural capital is poorly understood, scarcely monitored, and—in many cases—undergoing rapid degradation and depletion. To bring understanding of nature's values into decisions, the Natural Capital Project is developing models that quantify and map the values of ecosystem services. The modeling suite is best suited for analyses of multiple services and multiple objectives. The current models, which require relatively little data input, can identify areas where investment may enhance human well-being and nature. We are continuing to refine existing models and to develop new models.

We use the Millennium Ecosystem Assessment (2005) definition of the term ecosystem services: "the benefits people obtain from ecosystems." Ecosystems incorporate both biotic and abiotic components and we thus consider "ecosystem services" and "environmental services" to be equivalent. Natural capital is the living and non-living components of ecosystems that contribute to the provision of ecosystem services. Capital assets take many forms including manufactured capital (e.g., buildings and machines), human capital (knowledge, experience, and health), social capital (relationships and institutions), as well as natural capital.

2.2 Who should use InVEST?

InVEST is designed to inform decisions about natural resource management. Essentially, it provides information about how changes in ecosystems are likely to lead to changes in the flows of benefits to people. Decision-makers, from governments to non-profits to corporations, often manage lands and waters for multiple uses and inevitably must evaluate trade-offs among these uses. InVEST's multi-service, modular design provides an effective tool for exploring the likely outcomes of alternative management and climate scenarios and for evaluating trade-offs among sectors and services. For example, government agencies could use InVEST to help determine how to manage lands, coasts, and marine areas to provide a desirable range of benefits to people or to help design permitting and mitigation programs that sustain nature's benefits to society. Conservation organizations could use InVEST to better align their missions to protect biodiversity with activities that improve human livelihoods. Corporations, such as consumer goods companies, renewable energy companies, and water utilities, could also use InVEST to decide how and where to invest in natural capital to ensure that their supply chains are sustainable and secure.

InVEST can help answer questions like:

- Where do ecosystem services originate and where are they consumed?

- How does a proposed forestry management plan affect timber yields, biodiversity, water quality and recreation?
- What kinds of coastal management and fishery policies will yield the best returns for sustainable fisheries, shoreline protection and recreation?
- Which parts of a watershed provide the greatest carbon sequestration, biodiversity, and tourism values?
- Where would reforestation achieve the greatest downstream water quality benefits while maintaining or minimizing losses in water flows?
- How will climate change and population growth impact ecosystem services and biodiversity?
- What benefits does marine spatial planning provide to society in addition to food from fishing and aquaculture and secure locations for renewable energy facilities?

2.3 Introduction to InVEST

InVEST is a tool for exploring how changes in ecosystems are likely to lead to changes in benefits that flow to people.

InVEST often employs a production function approach to quantifying and valuing ecosystem services. A production function specifies the output of ecosystem services provided by the environment given its condition and processes. Once a production function is specified, we can quantify the impact of changes on land or in the water on changes on the level of ecosystem service output.

InVEST uses a simple framework delineating "supply, service, and value" to link production functions to the benefits provided to people (Figure 1).

"Supply" represents what is potentially available from the ecosystem (ie. what the ecosystem structure and function can provide). For example, this would be the wave attenuation and subsequent reduction in erosion and flooding onshore provided by a particular location and density of mangrove forest. "Service" incorporates demand and thus uses information about beneficiaries of that service (e.g., where people live, important cultural sites, infrastructure, etc.). "Value" includes social preference and allows for the calculation of economic and social metrics (e.g., avoided damages from erosion and flooding, numbers of people affected).

The InVEST toolset described in this guide includes models for quantifying, mapping, and valuing the benefits provided by terrestrial, freshwater, and marine systems. We group models in InVEST into three primary categories: 1) supporting services, 2) final services, and 3) tools to facilitate ecosystem service analyses. Supporting services underpin other ecosystem services, but do not directly provide benefits to people. Final services provide direct benefits to people. For final services, we split the services into their biophysical supply and the service to people wherever possible; these are delineated in the model names as Supply: Service. For some final services, we model the service directly, without modeling the supply separately. These are listed as the models below without the use of a colon, and also below in Figure 3.

2.3.1 Supporting Ecosystem Services:

- Marine Water Quality
- Habitat Risk Assessment
- Habitat Quality

2.3.2 Final Ecosystem Services:

- Carbon Storage and Sequestration: Climate Regulation
- Blue Carbon Storage and Sequestration: Climate Regulation

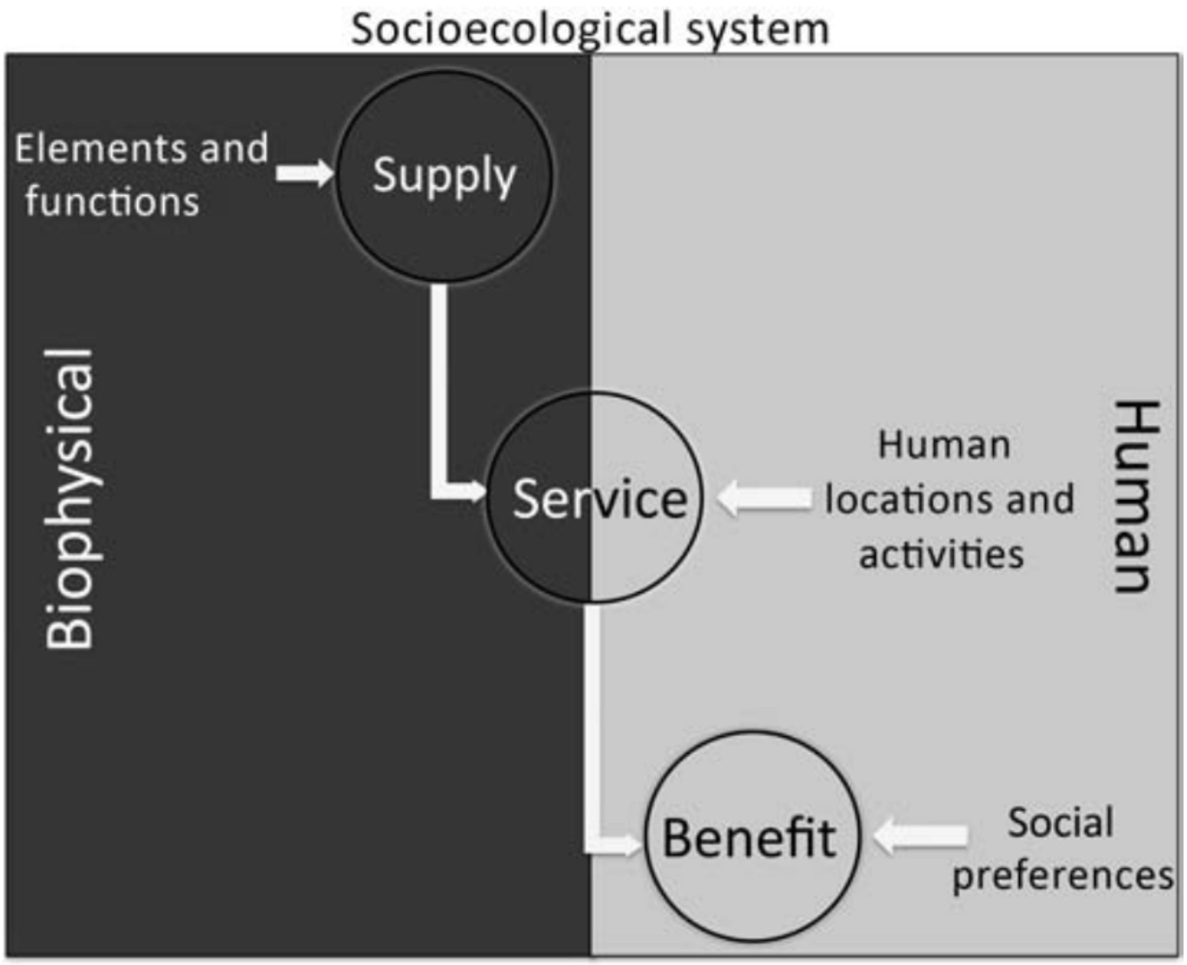


Fig. 2.1: The ecosystem service supply chain, linking ecological function to ecosystem services and the benefits provided to people (From Tallis et al. 2012, Bioscience)

- Water Yield: Reservoir Hydropower Production
- Nutrient Retention: Water Purification
- Sediment Retention: Avoided Dredging and Water Purification
- Pollinator Abundance: Crop Pollination
- Coastal Exposure and Vulnerability
- Wave Attenuation & Erosion Reduction: Coastal Protection (*only in ArcGIS version*)
- Unobstructed Views: Scenic Quality Provision
- Nature-based Recreation and Tourism
- Managed Timber Production
- Wave Energy Production
- Offshore Wind Energy Production
- Marine Finfish Aquacultural Production
- Marine Fisheries Production (*coming soon*)

2.3.3 Tools to Facilitate Ecosystem Service Analyses:

- RouteDEM (Hydrological routing functions)
- Overlap Analysis

2.4 Using InVEST to Inform Decisions

Information about changes in ecosystem services is most likely to make a difference when questions are driven by decision-makers and stakeholders, rather than by scientists and analysts. We have found that InVEST is most effective when used within a decision-making process. The Natural Capital Project has used InVEST in over 20 decision contexts worldwide (Figure 2). Through our experience applying InVEST and helping to shape decisions, we have seen how the InVEST tool fits within the larger context of a natural capital approach.

Our approach (Figure 4) starts with a series of stakeholder consultations. Through discussion, questions of interest to policy makers, communities and conservation groups are identified. These questions may concern service delivery on a landscape today and how these services may be affected by new programs, policies, and conditions in the future. For questions regarding the future, stakeholders develop “scenarios” to explore the consequences of expected changes on natural resources. These scenarios typically include a map of future land use and land cover or, for the marine models, a map of future coastal and ocean uses and coastal/marine habitats. These scenarios that are assessed for ecosystem service value by biophysical and economic models that produce several types of outputs. Following stakeholder consultations and scenario development, InVEST can estimate the amount and value of ecosystem services that are provided on the current landscape or under future scenarios. InVEST models are spatially explicit, using maps as information sources and producing maps as outputs. InVEST returns results in either biophysical terms, whether absolute quantities or relative magnitudes (e.g., tons of sediment retained or % of change in sediment retention) or economic terms (e.g., the avoided treatment cost of the water affected by that changed in sediment load; see Figure 3).

The spatial extent of analyses is also flexible, allowing users to address questions at the local, regional or global scale. InVEST results can be shared with the stakeholders and decision makers who created the scenarios to inform upcoming decisions. Using InVEST is an iterative process, and stakeholders may choose to create new scenarios based on the information revealed by the models until suitable solutions for management action are identified.

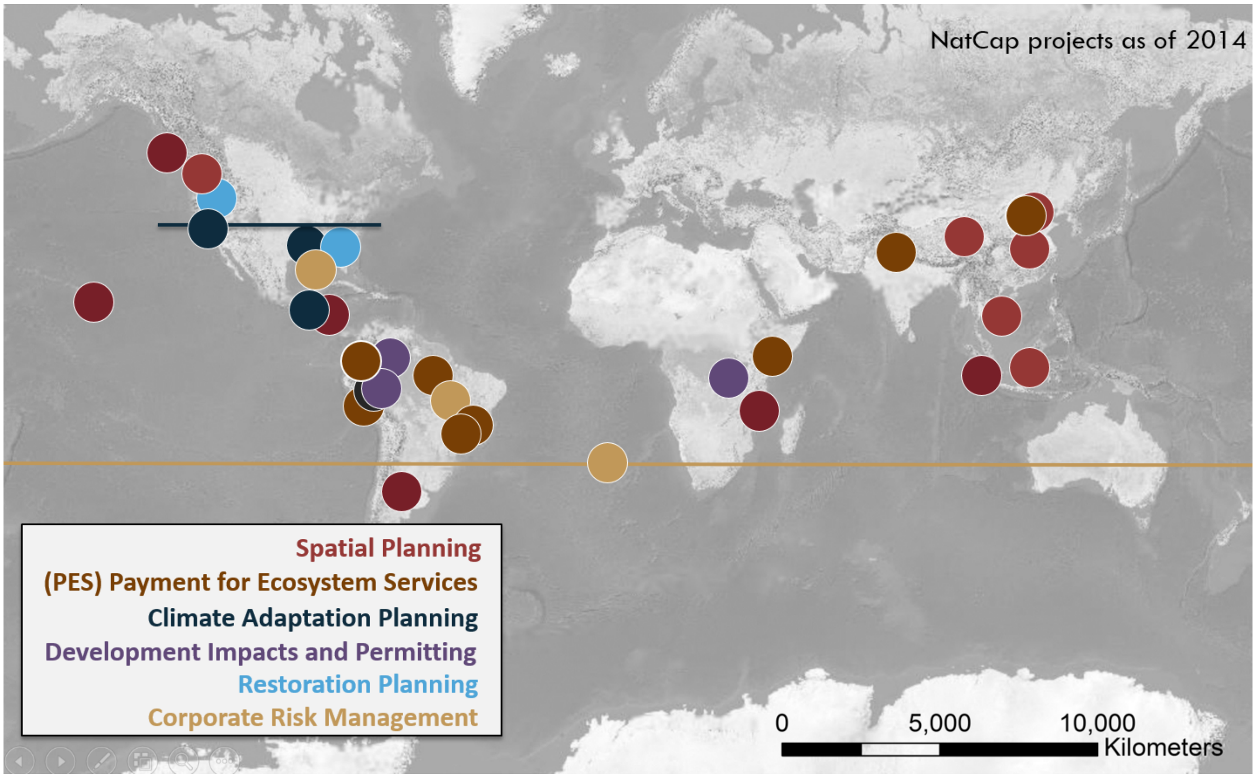


Fig. 2.2: The Natural Capital Project has worked to apply an ecosystem services-based approach across a wide range of decision contexts in over 20 locations around the world.

Figure 5 below provides some concrete examples of how the general approach can be used to inform different types of decisions.

2.5 A work in progress

InVEST is a free of cost software product licensed under the BSD open source license.

The development of InVEST is an ongoing effort of the Natural Capital Project. We release updated versions of the toolkit approximately every three months that can include updated science, performance and feature enhancements, bug fixes, and/or new models. As a historical note, the original InVEST models were built within ArcGIS but now almost all models exist in a standalone form directly launchable from the Windows Operating system with no other software dependencies. The exception is the "Wave Attenuation & Erosion Reduction: Coastal Protection" model which exists as an ArcGIS toolbox; this tool will be converted to the standalone platform.

A note on InVEST versioning: Integer changes will reflect major changes like the transition from 2.6.0 to 3.0.0 indicates a transition from the Arc-GIS modules to standalone version. An increment in the digit after the primary decimal indicates major new features (e.g, the addition of a new model) or major revisions. The third decimal reflects minor feature revisions or bug fixes with no new functionality.

2.6 This guide

This guide will help you understand the basics of the InVEST models and start using them. The next chapter leads you through the installation process and provides general information about the tool and interface.

	Biophysical/supply model	Service model	Spatial extent (given assumed input data resolution*)	Methods of valuation available: Biophysical amount	$**	Number of people affected	Forms of output possible: Maps	Relative estimates	Quantitative estimates
Supporting ecosystem services	Habitat quality		Any (with resolution finer than distance over which threats operate)				x	x	
	Habitat risk assessment		Any				x	x	
	Marine water quality		Any				x		x
Final ecosystem services	Carbon storage and sequestration	Climate regulation	Any	x	x		x		x
	Blue carbon (marine carbon storage and sequestration)	Climate regulation	Any	x	x		x		x
	Water yield	Reservoir hydropower production	Watersheds ranging from 1 sq km (if at least 10 m resolution) - global (for up to 1 km resolution)	x	x	x (with proper delineation of watersheds)	x	x (without calibration)	x (with calibration)
	Nutrient retention	Water purification	Watersheds ranging from 1 sq km (if at least 10 m resolution) - global (for up to 1 km resolution)	x	x	x (with proper delineation of watersheds)	x	x (without calibration)	x (with calibration)
	Sediment retention	Avoided dredging	Watersheds ranging from 1 sq km (if at least 10 m resolution) - global (for up to 1 km resolution)	x	x	x (with proper delineation of watersheds)	x	x (without calibration)	x (with calibration)
	Sediment retention	Water purification	Watersheds ranging from 1 sq km (if at least 10 m resolution) - global (for up to 1 km resolution)	x	x	x (with proper delineation of watersheds)	x	x (without calibration)	x (with calibration)
	Pollinator abundance	Crop pollination	Any (resolution must be less than pollinator foraging distance)				x	x	
	Potential protection from erosion and innundation (Coastal exposure)	Coastal protection screening tool (coastal vulnerability)	Any			x (with detailed population maps)	x	x	
	Wave attenuation & erosion reduction	Coastal protection	Local (most feasible over <50 mile extent)	x	x		x		x
	Unobstructed views	Scenic quality provision	Any			x (with detailed population maps)	x	x	x
		Nature-based recreation & tourism	Any	x			x	x	x (with calibration)
		Managed timber production	Any	x	x		x		x
		Wave energy production	Any, within global Exclusive Economic Zone	x	x		x	x	x
		Offshore wind energy production	Any, within global Exclusive Economic Zone	x	x		x	x	x
		Marine finfish aquacultural production	Any	x	x		x	x	x

*for most senstive inputs

**see model chapter for more information about monetary valuation assumptions

Fig. 2.3: The supporting and final ecosystem service models currently included in the InVEST software suite.

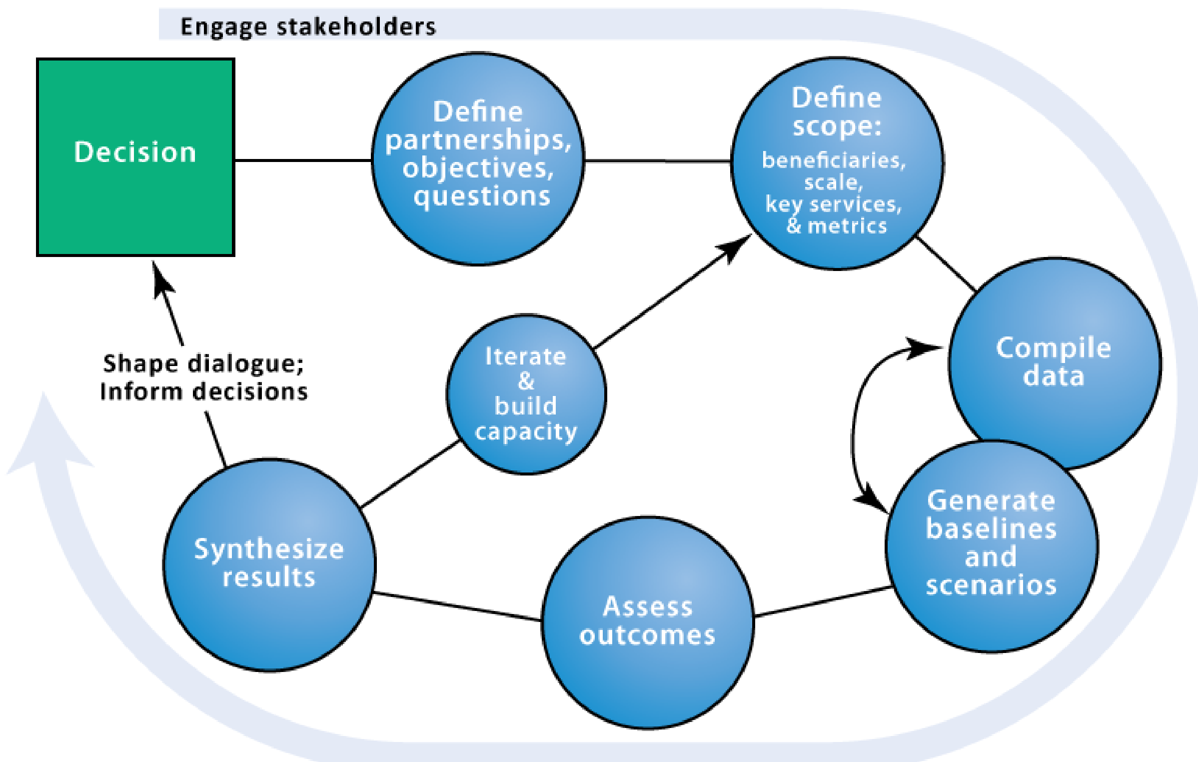


Fig. 2.4: Stages of a natural capital approach to informing decision making.

Decision context	Define partnerships, objectives, questions	Define scope: beneficiaries, scale, key services & metrics	Compile data	Generate baselines and scenarios	Assess outcomes	Synthesize results	Iterate & build capacity
PES design: Prioritze Water Fund allocations to conservation and restoration projects	Partners: NGOs, water fund participants. Objectives: Improve water quality and secure water supply through watershed conservation and restoration. Questions: Which activities should our fund invest in and where? How much return on our investment will we get?	Beneficiaries: Downstream city municipalities, agri-business, bottling companies; upstream farmers and ranchers. Scale: 50-1000 sq km watersheds with 30m resolution data. Key services & metrics: Kg sediment retained, cubic feet of water produced.	Data: Land-use, DEM, precipitation, soil; pre-processing to translate to erosivity, erodibility, PET, AET; stakeholder input on feasiblity, activity costs	Scenarios: Applying RIOS to select investment portfolio maps (identifying where to promote which activities) for different budget levels, to maximize water quality and quantity; applying to base land-cover to generate scenario maps	Assess outcomes: Run InVEST sediment retention and water yield models on base land-cover and investment scenario maps to evaluate the ecosystem services provided currently and by Water Funds investment	Synthesize results: Graphs or tables ofthe change in ecosystem service provided by current/business-as-usual versus	Iterate & build capacity: Co-develop a tailored tool with NGOs for building investment portfolios; Train water fund platform staff to provide technical support and trainings to others in the region
Infrastructure permitting: Evaluate proposed development project (e.g., mine, road) impacts and offsets	Partners: Local communities, government bodies, NGOs, industry. Objective: Minimize impacts and maximize benefits of development in a socially equitable way. Questions: What are the impacts of proposed development on ecosystem services and how are they distributed among communities?	Beneficiaries: Populations impacted by proposed development and/or mitigation options. Spatial extent: Watersheds or administrative units affected by development project. Services and metrics: For example, carbon sequestration and water-related services are likely to be impacted and important to mitigate, measured as % change in service.	Data: LULC, DEM, biophysical paramteters, soil data, climate data, human population data	Baseline and scenarios: Representing 1) the current landscape, 2) the landscape with development, and 3) the landscape with development and mitigation	Assess outcomes: Quantify climate regulation services (carbon storage) and drinking water quality (sediment and nutrient pollutant loads) across servicesheds before and after development, and with mitigation.	Synthesize results: Graphs and tables showing how much ecosystem services change with development and the amount of impact offset by mitigation. Maps and graphs shoing who is impacted, where the impacts are located and how large any ecosystem service losses are.	Iterate & build capacity: Consider effects of additional mitigation options, provide training to local partners
Corporate supply chain: identify sustainable sourcing strategies	Partners: procurement or R&D divisions of consumer goods corporations. Objective: identify sourcing regions and/or management practices for more sustainable production. Questions: what regions will be able to meet increased demand with minimal impacts on biodiversity and ecosystem services? what management practices should be prioritzed in different sourcing regions?	Beneficiaries: Global population, consumers, and/or populations living within sourcing regions. Spatial extent: Ranges, but often municipalities or larger. Services and metrics: Biodiversity, carbon, water, often measured as (relative or absolute) change associated with different sourcing strategies.	Data: LULC, DEM, biophysical paramteters, soil data, climate data, human population data	Baseline and scenarios: Considering how to translate a change in commodity demand to a concomitant change in land-use or land management; or representing different best management practices that could be applied by producers in supply chain	Assess outcomes: Quantify change in biodiversity, carbon storage, and water quality and/or quantity for different sourcing strategies	Synthesize results: Maps of different production patterns and accompanying graphs comparing impacts among different sourcing strategies (regions, assumed patterns of land use/management change)	Iterate & build capacity: Develop detailed guidance and model scripts to facilitate independent runs by corporate sustainability teams to be able to evaluate future soucing decisions in-house
Spatial planning: create a marine spatial plan with zones of use for various activities	Partner: Central agency in charge, clear understanding of participating sectors/jurisdictions Objective: Sustainable delivery of full range of benefits Questions: How can we rationally arrange activities along coast? How might one	Spatial extent: Exclusive economic zone (EEZ), or one bay + watershed. Key services/metrics: Fisheries landings (lbs,$), shoreline protection (area, $), tourism (visitors, $)	Data: Coastline, LULC, bathymetry, coastal habitats, infrastructure, landings, wave/wind	Baseline: Current uses of marine space + current LULC, habitats, etc. Scenarios: Possible future uses of marine space + effect on habitats (can be through habitat risk assessment model)	Assess outcomes: Explore cumulative impacts of activities on habitats, compare landings, storm damages, tourism rates across alternative management schemes	Synthesize results: Create spider diagram showing trade-offs, make maps that highlight differences across management schemes, make bar charts	Iterate & build capacity: Get feedback on results, create new scenarios, train agency staff to run models
Climate adaptation: design and implement a sea-level rise adaptation strategy	Partners: Central agency in charge, clear understanding of participating sectors/jurisdictions Objective: Cost-effectively protect people and property while maximizing co-benefits Questions: Can green infrastructure provide adequate protection? What else does it provide?	Spatial extent: A particular stretch of coastline (e.g. county,bay) Key services/metrics: Shoreline protection (area, $), tourism (visitors, $), carbon sequestered (mT, $) landings (lbs, $)	Data: Coastline, LULC, bathymetry, coastal habitats, infrastructure, landings, wave/wind	Baseline: Current LULC, habitats, seawalls, levees etc. Scenarios: SLR, habitat migration (e.g. from SLAMM), possible adaptation alternatives	Assess outcomes: Compare storm damages, carbon storage, landings, tourism rates across alternative climate/habitat scenarios	Synthesize results: Compare protection provided by alternative scenarios (area, $, people), explore how green infrastructure + grey infrastructure work together, assess /compare co-benefits	Iterate & build capacity: Get feedback on results, create new scenarios, train agency staff to run models

Fig. 2.5: Examples of how the Natural Capital Project has used an ecosystem services approach to inform decisions across a variety of contexts. The columns in this table map onto the stages of the natural capital approach illustrated in Figure 3 above.

The remaining chapters present the ecosystem service models. Each chapter:

- briefly introduces a service and suggests the possible uses for InVEST results;
- explains intuitively how the model works, including important simplifications, assumptions, and limitations;
- describes the data needed to run the model, which is crucial because the meaning and value of InVEST results depend on the input data;
- provides step-by-step instructions for how to input data and interact with the tool;
- offers guidance on interpreting InVEST results;
- includes an appendix of information on relevant data sources and data preparation advice (this section is variable among chapters, and will improve over time from user input).

Much of the theory related to the scientific foundation of many of these models can be found in the book Natural Capital: The Theory & Practice of Mapping Ecosystem Services (Oxford University Press). The models applied and discussed in that book are not identical to those presented in the InVEST toolset, however, and this user guide provides the most up-to-date description of the current versions of the models. .. primerend

CHAPTER

THREE

GETTING STARTED

3.1 Installing InVEST and sample data on your computer

Download the InVEST installer from www.naturalcapitalproject.org. The executable will be called "InVEST_<version>_Setup.exe". Double-click on this .exe to run the installer.

After clicking through the first screen and agreeing to the Licence Agreement, the Choose Components screen will appear. The installer will always install the InVEST Tools, ArcGIS toolbox and HTML and PDF versions of the InVEST User's Guide. Optionally, sample datasets may also be installed, and by default they are all selected. Note that these datasets are downloaded over the internet, and some are very large (particularly the Marine Datasets), so they make take a long time to install. If you do not wish to install all or some of the sample datasets, uncheck the corresponding box(es).

Next, choose the folder where the InVEST toolsets and sample data will be installed. The installer shows how much space is available on the selected drive. Click Install to begin the installation.

Once installed, the InVEST install folder will contain the following:

- A **documentation** folder, containing the InVEST User Guide in HTML format.
- An **invest_helper_utils** folder, containing several ArcGIS tools to help with pre- and post-processing of InVEST data
- An **invest-3-x86** folder, containing the compiled Python code that makes up the InVEST toolset.
- A **python** folder, containing the Python scripts for the Coastal Protection model
- **coastal_protection_t1.tbx** the Marine Coastal Protection model, which is the last remaining InVEST model in ArcGIS format.
- **InVEST_<version>_Documentation.pdf**, the InVEST User Guide in PDF format.
- **Uninstall_<version>.exe**, which will uninstall InVEST.
- **Updates_InVEST_<version>.txt**, lists of all of the updates included in each new version.

Additionally, shortcuts for all InVEST standalone applications will be added to your Windows start menu under *All Programs -> InVEST +VERSION+*

3.1.1 Standalone InVEST Tools

All but one of the InVEST models run on an entirely open-source platform, where historically the toolset was a collection of ArcGIS scripts. The new interface does not require ArcGIS and the results can be explored with any GIS tool including ArcGIS, QGIS, and others. As of InVEST 2.3.0, the toolset has had standalone versions of the models available from the Windows start menu after installation, under *All Programs -> InVEST +VERSION+*. Standalone versions are currently available for all models except Coastal Protection. The ArcGIS versions of all other models are no longer supported.

** *The following directions apply only to the remaining ArcGIS model, Coastal Protection.* *

3.1.2 Getting started with the InVEST Coastal Protection model in ArcGIS

The InVEST Coastal Protection model runs as a script tool in the ArcGIS ARCTOOLBOX environment. To run this model, you must have:

- ArcGIS 9.3 (service pack 1 or 2) or ArcGIS 10 (service pack 1).
- ArcINFO level license to run some of the models
- Spatial Analyst extension installed & activated
- Additional Python libraries available for download at www.naturalcapitalproject.org

Running InVEST in ArcGIS does not require Python programming, but it does require basic to intermediate skills in ArcGIS.

Downloading and installing Python library extensions

Users running the InVEST Coastal Protection model are required to download the Python extensions file found on the InVEST installer download page at www.naturalcapitalproject.org.

Coastal Protection requires the following extensions to be installed:

1. Numeric Python (**NumPy**) is a powerful and flexible N-dimensional array container that provides the fundamentals needed for scientific computing in Python. An older incompatable version of NumPy comes standard with the ArcGIS 9.3 and 10. While the "Marine Python Extension Check" tool will confirm that NumPy is already installed on your machine, make certain to install the latest version of NumPy from the InVEST installer download page.
2. Scientific Library for Python (**SciPy**) is an Open Source library of scientific tools for Python. It calls on the NumPy library and gathers a variety of high-level science and engineering modules together as a single package.
3. Python for Windows (**PythonWin**) allows users to access data from Windows applications like Microsoft Excel.
4. **Matplotlib** is a Python 2D plotting library which produces publication quality figures.

See the Marine InVEST FAQ for help with installing these extensions.

Adding the Coastal Protection toolbox to ArcGIS

- START ArcGIS. Save as a new mxd file. Ensure that ArcToolbox is open. If not, select the toolbox icon from the standard toolbar.
- Right-click on an empty part of the ArcToolbox window and select ADD TOOLBOX. Or, right click on the top-most ArcToolbox text.
- Navigate to the location of coastal_protection_t1.tbx, in the InVEST folder. Select the toolbox and click OPEN. Do not double click on the toolbox icon.
- The Coastal Protection toolbox should appear in ArcToolbox. Click on the plus sign to the left of Coastal Protection Tier 1 to expand it. You will see the three scripts that make up the Coastal Protection model.

3.2 Using sample data

InVEST comes with sample data as a guide for formatting your data. For instance, in preparation for analysis of your data, you may wish to test the models by changing input values in the sample data to see how the output responds.

Sample data are found in separate thematic folders in the InVEST folder. For example, the sample datasets for the Pollination model are found in \{InVEST install directory}\pollination\input, and those for the Carbon model in \{InVEST install directory}\carbon\input. When opening the models, you'll notice that default paths point to these sample datasets. You will also notice that the default workspace for each tool is the thematic folder with a name that matches the tool. Once you are working with your own data, you will need to create a workspace and input data folders to hold your own input and results. You will also need to redirect the tool to access your data.

3.3 Formatting your data

Before running InVEST, it is necessary to format your data. Although subsequent chapters of this guide describe how to prepare input data for each model, there are several formatting guidelines common to all models:

- Data file names should not have spaces (e.g., a raster file should be named 'landuse' rather than 'land use').
- If using ESRI GRID format rasters, their dataset names cannot be longer than 13 characters and the first character cannot be a number. TIFF and IMG rasters do not have the length limitation.
- Spatial data should be projected, and all input data for a given tool should be in the same projection. If your data is not projected InVEST will often give incorrect results.
- While the InVEST 3.0 models are now very memory-efficient, the amount of time that it takes to run the models is still affected by the size of the input datasets. If the area of interest is large and/or uses rasters with small cell size, this will increase both the memory usage and time that it takes to run the model. If they are too large, a memory error will occur. If this happens, try reducing the size of your area of interest, or using coarser-resolution input data.
- For Coastal Protection, results will be calculated on selections in tables and feature classes. If you are setting the model to read layers and tables from your ArcGIS document rather than from the c-drive, make sure to clear any selections unless you wish to run your model on the selection.
- Running the models with the input data files open in another program can cause errors. Ensure that the data files are not in use by another program to prevent data locking.
- Regional and Language options: Some language settings cause errors while running the models. For example settings which use coma (,) for decimals instead of period (.) cause errors in the models. To solve this change the regional settings to English.
- As the models are run, it may be necessary to change values in the input tables. This can happen within ArcGIS or in an external program. Depending on the format of tables used (dbf or mdb is recommended) you will need an appropriate software program to edit tables. To edit tables within ArcGIS, you need to start an edit session (from the editor toolbar) and select the workspace (folder or database) that contains your data. After editing you must save your changes and stop the edit session.
- Some models require specific naming guidelines for data files (e.g., Biodiversity model) and field (column) names, which are defined in the User Guide chapter for each model. Follow these carefully to ensure your dataset is valid.
- Remember to use the sample datasets as a guide to format your data.

3.4 Running the models

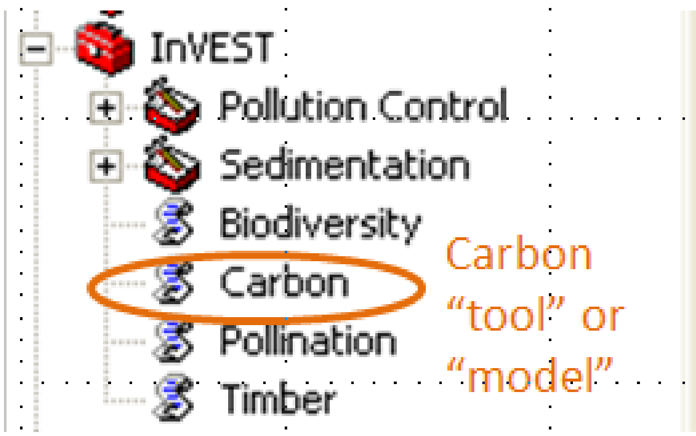
You are ready to run an InVEST model when you have prepared your data according to the instructions in the relevant chapter, installed InVEST, and, if using the Coastal Protection model, adding its toolbox to your ArcGIS document.

To begin:

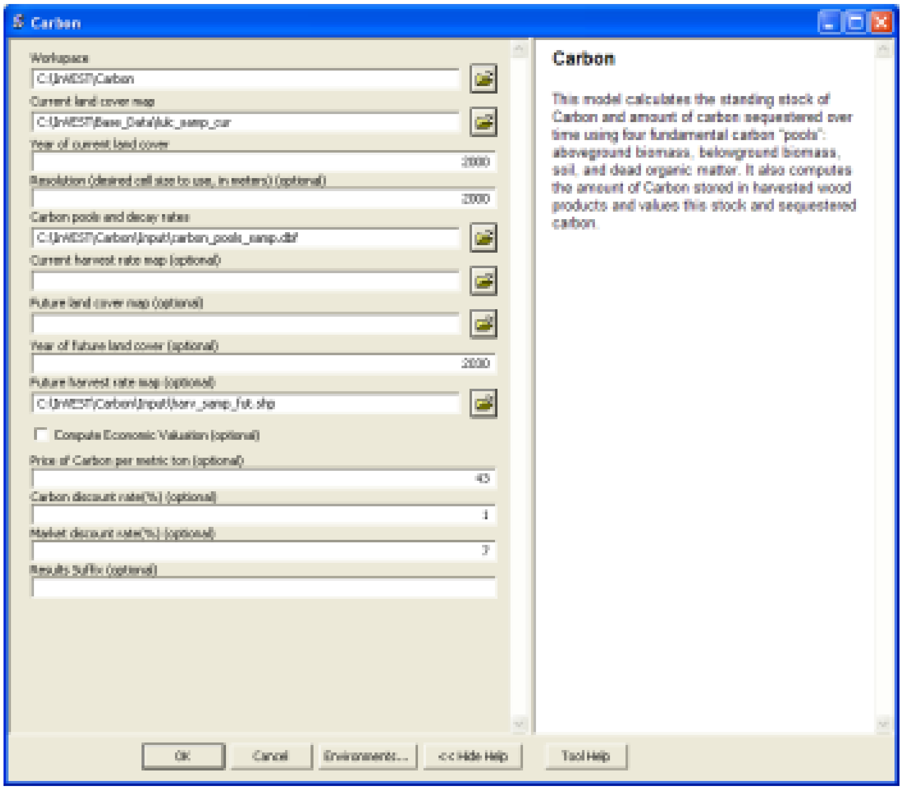
- Although not necessary, it's often useful to add your input layers to your GIS document to examine them. Use the ADD DATA button to add input data for a given module.
- View the attribute table by right-clicking on the layer and selecting OPEN ATTRIBUTE TABLE. You can change the symbology of an input layer by right-clicking on the layer name in the TABLE OF CONTENTS and selecting PROPERTIES, then clicking on the SYMBOLOGY tab.

Note: Some of the models make changes to the data tables as they run. Such models will not run correctly if the tables are added to the map as the data will be locked.

- Double-click the model you wish to run (e.g., Carbon) and complete the required parameters in the dialogue box that appears.



- The Carbon dialog is shown below as an example. Fields for which the entered path leads to a non-existent file will be marked with a red "x" next to the space for that variable. You can run the model with sample data as shown by the default paths, or navigate the paths to your data. Instructions specific for each model are in subsequent chapters.
- Note that each tool has a place to enter a suffix to the output filenames. Adding a unique suffix prevents overwriting files produced in previous iterations. When all required fields are filled in, click the OK button on the interface.
- Processing time will vary depending on the script and the resolution and the extent of the datasets in the analysis. Every model will open a window showing the progress of the script. Be sure to scan the output window for useful messages. Normal progress notes will be printed in black font. Informative messages that may or may not require changes to the data will be indicated in green font. Messages in red font indicate problems that have caused the model not to run. Read the green and red messages carefully to be aware of potential data problems or to determine why the model did not produce an output.
- The model creates two folders in the workspace you selected: 'intermediate' and 'output.' After your script completes successfully, you can view the results by adding them from the folders to your ArcGIS document using the ADD DATA button. View the attribute table and change SYMBOLOGY, by right-clicking on the layer name in the TABLE OF CONTENTS and selecting PROPERTIES, then clicking on the SYMBOLOGY tab.



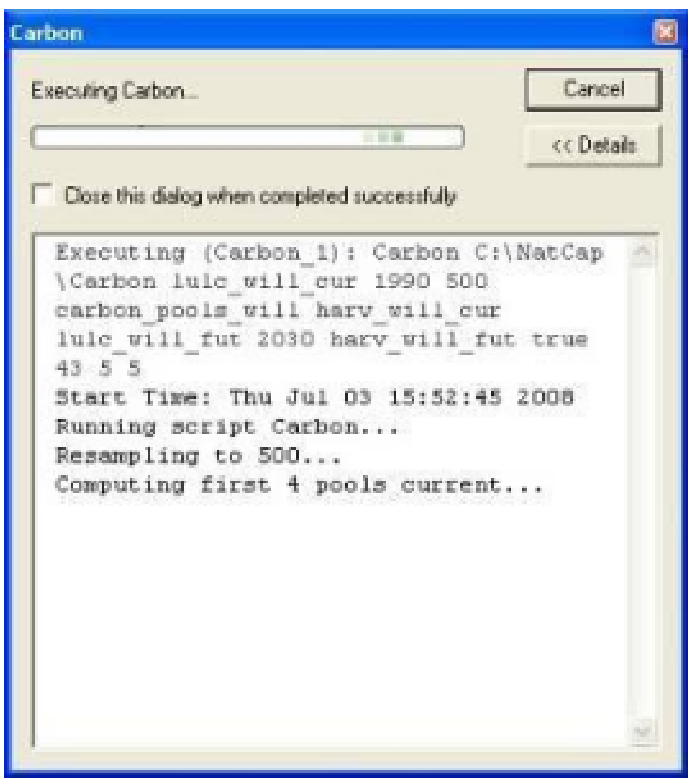


Fig. 3.1: Progress dialog

3.5 Support information

Several regular training workshops on InVEST may be offered annually, subject to funding and demand. Information on these trainings will be announced on the support page and can be found at the Natural Capital Project website. This site is also a good source of general information on InVEST and other activities of the Natural Capital Project.

3.6 Reporting errors

If you encounter any issues please post to the user's support forum at http://forums.naturalcapitalproject.org with the following information:

- InVEST model you're having difficulty with
- Explicit error message or behavior
- If possible, a screenshot of the state of your InVEST toolset when you get the error.
- The log file produced by the standalone models, located in the output Workspace folder.

3.7 Working with the DEM

For the hydrology tools Nutrient Retention and Sediment, having a well-prepared digital elevation model (DEM) is critical. It must have no missing data and should correctly represent the surface water flow patterns over the area of interest in order to get accurate results.

Here are some tips for working with the DEM and creating a hydrologically-correct DEM. Included is information on using built-in ArcGIS Spatial Analyst functions as well as ArcHydro (see resources below), an ArcGIS data model that has a more complex and comprehensive set of tools for modeling surface water features. ArcSWAT, AGWA, and BASINS, which are not covered here, are other options for delineating watersheds and doing watershed processing. This is only intended to be a brief overview of the issues and methods involved in DEM preparation. For more detail, see the Resources section below.

- Use the highest quality, finest resolution DEM that is appropriate for your application. This will reduce the chances of there being sinks and missing data, and will more accurately represent the terrain's surface water flow, providing the amount of detail that is required for making informed decisions at your scale of interest.
- Mosaic tiled DEM data

  If you have downloaded DEM data for your area that is in multiple, adjacent tiles, they will need to first be mosaicked together to create a single DEM file. In ArcToolbox, use Data Management -> Raster -> Mosaic to New Raster, entering all of the tiles into the Input Rasters list. Look closely at the output raster to make sure that the values are correct along the edges where the tiles were joined. If they are not, try different values for the Mosaic Method parameter to the Mosaic to New Raster tool.
- Check for missing data

  After getting (and possibly mosaicking) the DEM, make sure that there is no missing data (or 'holes'), represented by NoData cells within the area of interest. If there are NoData cells, they must be assigned values.

  For small holes, one way to do this is to use the ArcGIS Focal Mean function within Raster Calculator (or Conditional -> CON). For example, in ArcGIS 9.3.x:

```
con(isnull([theDEM]), focalmean([theDEM], rectangle, 3, 3), [theDEM])
```

  In ArcGIS 10.x:

```
Con(IsNull("theDEM"),FocalStatistics("theDEM",NbrRectangle(3,3),"MEAN"),"theDEM")
```

Interpolation can also be used, and can work better for larger holes. Convert the DEM to points using Conversion Tools -> From Raster -> Raster to Point, interpolate using Spatial Analyst's Interpolation tools, then use CON to assign interpolated values to the original DEM:

```
con(isnull([theDEM]), [interpolated_grid], [theDEM])
```

Another possibility is assigning data from a different DEM, if surrounding values are a good match, again using CON:

```
con(isnull([theDEM]), [different_DEM], [theDEM])
```

- Verify the stream network

  The stream network generated by the model from the DEM should closely match the streams on a known correct stream map. Several of the InVEST hydrology models, as well as the pre-processing tool for RIOS output the stream network generated by the tool, corresponding to the 'threshold flow accumulation' value input to the model. Use these outputs to evaluate how well the modelled streams match reality, and adjust the threshold flow accumulation accordingly. Larger values of the threshold will produce coarser stream networks with fewer tributaries, smaller values will produce more tributaries.

  Or, create these stream layers manually using the following steps in ArcGIS:

  1. Generate a flow accumulation map from the DEM using Hydrology -> Flow Direction then Hydrology -> Flow Accumulation.
  2. Create the stream network with the tool Math -> Logical -> Greater Than Equal, with the flow accumulation raster as Input raster 1 and the threshold flow accumulation value as Input raster 2. Compare the resulting stream layer to a known correct stream map. Repeat, adjusting the threshold value, until the resulting streams most closely match.

  If the generated stream network does not look correct, continue with the following steps in ArcGIS to 'burn' the correct stream network into the DEM. Note that this is a very simplistic way of burning in streams, and there are other, more complex methods that may produce better results.

  1. If starting with a vector stream layer that is known to be correct, convert it to a grid that has the same cell size and extent as the DEM.
  2. Assign the stream grid a cell value of 1 where there are streams and 0 elsewhere.
  3. Subtract a multiple of this stream grid from the DEM.

  If using ArcHydro, create the stream network from the DEM using Terrain Preprocessing -> Stream Definition and compare it to a known correct stream layer. If the generated stream network does not look correct, 'burn' the correct stream layer in using the Terrain Preprocessing -> DEM Manipulation -> DEM Reconditioning function.

- Identify sinks in the DEM and fill them

  From the ESRI help on "How Sink works": "A sink is a cell or set of spatially connected cells whose flow direction cannot be assigned one of the eight valid values in a flow direction raster. This can occur when all neighboring cells are higher than the processing cell or when two cells flow into each other, creating a two-cell loop."

  Sinks are usually caused by errors in the DEM, and they can produce an incorrect flow direction raster. Possible by-products of this are areas with circular flow direction (or a 'loop') or a discontinuous flow network. Filling the sinks assigns new values to the anomalous processing cells, such that they are better aligned with their neighbors. But this process may create new sinks, so an iterative process may be required.

  In ArcGIS, first identify sinks using the Hydrology -> Sink tool. Fill the resulting sinks with Hydrology -> Fill. Do further iterations if there are still sinks that need to be filled.

In ArcHydro, the corresponding tools are Terrain Preprocessing -> DEM Manipulation -> Sink Evaluation and Fill Sinks.

- A note about reprojecting DEMs

  When reprojecting a DEM in ArcGIS, it is important to select BILINEAR or CUBIC for the "Resampling Technique." Selecting NEAREST will generally produce a DEM with an incorrect grid pattern across the area of interest, which might only be obvious when zoomed-in or after Flow Direction has been run.

- Creating watersheds

  To create watersheds in ArcGIS, it may be possible to use the Hydrology -> Watershed tool, which requires an input flow direction grid (created from the DEM using the Flow Direction tool) and point data for the locations of your points of interest (which represent watershed outlets, reservoirs, hydropower stations etc), snapped to the nearest stream using the Snap Pour Point tool. If the modeled watersheds are too large or too small, go back to the Snap Pour Point step and choose a different snapping distance or try an alternate method of delineation.

  In ArcHydro, there is a more lengthy process, which tends to produce more reliable results than the Watershed tool. Use the Watershed Processing -> Batch Watershed Delineation tool, which requires the creation of a flow direction grid, streams, catchments and point data for the locations of your points of interest, all done within the ArcHydro environment. See the ArcHydro documentation for more information.

  After watersheds are generated, verify that they represent the catchments correctly and that each watershed is assigned a unique integer ID in the field "ws_id"

- Creating sub-watersheds

  Sub-watersheds are now required for all of the InVEST hydrology models. For the Water Purification and Sediment models in ArcGIS, each sub-watershed must be smaller than the equivalent of approximately 4000 x 4000 pixels, due to limitations with Python and the ArcGIS memory model. (This limit does not apply to the standalone versions of these models.)

  To create sub-watersheds in ArcGIS, use the Hydrology -> Watershed tool. In this case, the input point data will represent multiple points along the stream network within the main watershed, such that a sub-watershed will be generated for each.

  In ArcHydro, use the Watershed Processing -> Batch Subwatershed Delineation tool, with input point data representing multiple points along the stream network within the main watershed. A sub-watershed will be generated for each point.

  Again, after the sub-watersheds are generated, verify that they represent the catchments correctly. Ensure each sub-watershed is assigned a unique integer ID in the field "subws_id" and that no duplicates are present.

3.8 Resources

ArcHydro: http://resources.arcgis.com/en/communities/hydro/01vn00000010000000.htm

ArcSWAT: http://swatmodel.tamu.edu/software/arcswat

AGWA: http://www.epa.gov/esd/land-sci/agwa/

BASINS: http://water.epa.gov/scitech/datait/models/basins/index.cfm

For more information on and an alternate method for creating hydrologically correct surfaces, see the ESRI help on "Hydrologically Correct Surfaces (Topo to Raster)".

For more information on sinks, see the ESRI help on "Creating a depressionless DEM".

Much more information and tips for all of these processes can be found by searching the ESRI support website.

Part II

InVEST Models

CHAPTER

FOUR

SUPPORTING ECOSYSTEM SERVICES:

4.1 Habitat Quality

4.1.1 Summary

Biodiversity is intimately linked to the production of ecosystem services. Patterns in biodiversity are inherently spatial, and as such, can be estimated by analyzing maps of land use and land cover (LULC) in conjunction with threats. InVEST models habitat quality and rarity as proxies for biodiversity, ultimately estimating the extent of habitat and vegetation types across a landscape, and their state of degradation. Habitat quality and rarity are a function of four factors: each threat's relative impact, the relative sensitivity of each habitat type to each threat, the distance between habitats and sources of threats, and the degree to which the land is legally protected. Required inputs include a LULC map, the sensitivity of LULC types to each threat, spatial data on the distribution and intensity of each threat and the location of protected areas. The model assumes that the legal protection of land is effective and that all threats to a landscape are additive.

4.1.2 Introduction

A primary goal of conservation is the protection of biodiversity, including the range of genes, species, populations, habitats, and ecosystems in an area of interest. While some consider biodiversity to be an ecosystem service, here we treat it as an independent attribute of natural systems, with its own intrinsic value (we do NOT monetize biodiversity in this model). Natural resource managers, corporations and conservation organizations are becoming increasingly interested in understanding how and where biodiversity and ecosystem services align in space and how management actions affect both.

Evidence from many sources builds an overwhelming picture of pervasive biodiversity decline worldwide (e.g., Vitousek et al. 1997; Wilcove et al 1998; Czech et. al 2000). This evidence has prompted a wide-ranging response from both governments and civil society. Through the Rio Convention on Biodiversity, 189 nations have committed themselves to preserving the biodiversity within their borders. Yet, there is scant research on the overlap between opportunities to protect biodiversity and to sustain the ecosystem services so critical to these countries' economic well-being. This is precisely the type of challenge that InVEST has been designed to address.

For managers to understand the patterns of distribution and richness across a landscape, individually and in aggregate, it is necessary to map the range or occurrences of elements (e.g. species, communities, habitats). The degree to which current land use and management affects the persistence of these elements must also be assessed in order to design appropriate conservation strategies and encourage resource management that maximizes biodiversity in those areas.

There are a variety of approaches to identifying priorities for conservation with various trade-offs among them. Each of these approaches focuses on different facets of biodiversity attributes and dynamics, including habitat or vegetation-based representation (i.e., a coarse filter), maximizing the number of species "covered" by a network of conserved sites for a given conservation budget (Ando et al. 1998), identifying patterns of richness and endemism (CI hotspots), and conserving ecological processes. There is also a hybrid coarse-fine filter approach which selectively includes "fine-filter" elements such as species with unique habitat requirements who may not be adequately protected using a coarse-filter approach only (TNC and WWF ecoregional planning). The InVEST Habitat Quality model is most relevant to "coarse filter", or habitat-based approaches.

The reasons for modeling biodiversity alongside ecosystem services are simple and powerful. Doing so allows us to compare spatial patterns of biodiversity and ecosystem services, and to identify win-win areas (i.e., areas where conservation can benefit both natural systems and human economies) as well as areas where these goals are not aligned. Further, it allows us to analyze trade-offs between biodiversity and ecosystem services across differing scenarios of future land use change. Land use/land cover (LULC) patterns that generate greater ecosystem service production may not always lead to greater biodiversity conservation (Nelson et al. 2008), and modeling future options today can help identify and avoid tradeoffs.

4.1.3 The Model

The InVEST habitat quality model combines information on LULC and threats to biodiversity to produce habitat quality maps. This approach generates two key sets of information that are useful in making an initial assessment of conservation needs: the relative extent and degradation of different types of habitat types in a region and changes across time. This approach further allows rapid assessment of the status of and change in a proxy for more detailed measures of biodiversity status. If habitat changes are taken as representative of genetic, species, or ecosystem changes, the user is assuming that areas with high quality habitat will better support all levels of biodiversity and that decreases in habitat extent and quality over time means a decline in biodiversity persistence, resilience, breadth and depth in the area of decline.

The habitat rarity model indicates the extent and pattern of natural land cover types on the current or a potential future landscape vis-a-vis the extent of the same natural land cover types in some baseline period. Rarity maps allow users to create a map of the rarest habitats on the landscape relative to the baseline chosen by the user to represent the mix of habitats on the landscape that is most appropriate for the study area's native biodiversity.

The model requires basic data that are available virtually everywhere in the world, making it useful in areas for which species distribution data are poor or lacking altogether. Extensive occurrence (presence/absence) data may be available in many places for current conditions. However, modeling the change in occurrence, persistence, or vulnerability of multiple species under future conditions is often impossible or infeasible. While a habitat approach leaves out the detailed species occurrence data available for current conditions, several of its components represent advances in functionality over many existing biodiversity conservation planning tools. The most significant is the ability to characterize the sensitivity of habitats types to various threats. Not all habitats are affected by all threats in the same way, and the InVEST model accounts for this variability. Further, the model allows users to estimate the relative impact of one threat over another so that threats that are more damaging to biodiversity persistence on the landscape can be represented as such. For example, grassland could be particularly sensitive to threats generated by urban areas yet moderately sensitive to threats generated by roads. In addition, the distance over which a threat will degrade natural systems can be incorporated into the model.

Model assessment of the current landscape can be used as an input to a coarse-filter assessment of current conservation needs and opportunities. Model assessment of potential LULC futures can be used to measure potential changes in habitat extent, quality, and rarity on a landscape and conservation needs and opportunities in the future.

How it Works

Habitat Quality

We define habitat as "the resources and conditions present in an area that produce occupancy – including survival and reproduction – by a given organism (Hall et al. 1997:175)." Habitat quality refers to the ability of the ecosystem to provide conditions appropriate for individual and population persistence, and is considered a continuous variable in the model, ranging from low to medium to high, based on resources available for survival, reproduction, and population persistence, respectively (Hall et al 1997). Habitat with high quality is relatively intact and has the structure and function within the range of historic variability. Habitat quality depends on a habitat's proximity to human land uses and the intensity of these land uses. Generally, habitat quality is degraded as the intensity of nearby land-use increases (Nelleman 2001, McKinney 2002, Forman et al. 2003).

The model runs using raster data, or a gridded map of square cells. Each cell in the raster is assigned a LULC type, which can be a natural (unmanaged) cover or a managed cover. LULC types can be given at any level of classification detail. For example, grassland is a broad LULC definition that can be subdivided into pasture, restored prairie, and residential lawn types to provide much more LULC classification detail. While the user can submit up to 3 raster maps of LULC, one each for a baseline, current, and future period, at a minimum the current LULC raster map has to be submitted.

The user defines which LULC types can provide habitat for the conservation objective (e.g., if forest breeding birds are the conservation objective then forests are habitat and non-forest covers are not habitat). Let H_j indicate the habitat suitability of LULC type j.

Which LULC types should be considered habitat? If considering biodiversity generally or if data on specific biodiversity-habitat relationships are lacking, you can take a simple binary approach to assigning habitat to LULC types. A classic example would be to follow an island-ocean model and assume that the managed land matrix surrounding remnant patches of unmanaged land is unusable from the standpoint of species (e.g., MacArthur and Wilson 1967). In this case a 0 would be assigned to managed LULC types in the matrix (i.e., non-habitat) and a 1 to unmanaged types (i.e., habitat). Under this modeling scheme habitat quality scores are not a function of habitat importance, rarity, or suitability; all habitat types are treated equally. Model inputs are assumed to not be specific to any particular species or species guild, but rather apply to biodiversity generally.

More recent research suggests that the matrix of managed land that surrounds patches of unmanaged land can significantly influence the "effective isolation" of habitat patches, rendering them more or less isolated than simple distance or classic models would indicate (Ricketts 2001, Prugh et al. 2008). Modification of the matrix may provide opportunities for reducing patch isolation and thus the extinction risk of populations in fragmented landscapes (Franklin and Lindenmayer 2009). To model this, a relative habitat suitability score can be assigned to a LULC type ranging from 0 to 1 where 1 indicates the highest habitat suitability. A ranking of less than 1 indicates habitat where a species or functional group may have lower survivability. Applying this second approach greatly expands the definition of habitat from the simple and often artificial binary approach (e.g., "natural" versus "unnatural") to include a broad spectrum of both managed and unmanaged LULC types. By using a continuum of habitat suitability across LULC types, the user can assess the importance of land use management on habitat quality holistically or consider the potential importance of "working" (or managed) landscapes.

If a continuum of habitat suitability is relevant, weights with a roster of LULC on a landscape must be applied in reference to a particular species guild of group. For example, grassland songbirds may prefer a native prairie habitat above all other habitat types (the habitat score for the LULC prairie (Hprarie) equals 1), but will also make use of a managed hayfield or pasture in a pinch (the habitat score for the LULC hayfield (Hhayfield) and pasture (Hpasture) equals 0.5). However, mammals such as porcupines will find prairie unsuitable for breeding and feeding. Therefore, if specific data on species group-habitat relationships are used, the model output refers to habitat extent and quality for the species or group in the modeled set only.

Besides a map of LULC and data that relates LULC to habitat suitability, the model also requires data on habitat threat density and its affects on habitat quality. In general, we consider human modified LULC types that cause habitat fragmentation, edge, and degradation in neighboring habitat threats. For example, the conversion of a habitat LULC to non-habitat LULC reduces the size and continuity of neighboring habitat patches. Edge effects refer to changes in the biological and physical conditions that occur at a patch boundary and within adjacent patches. For example, adjacent degraded non-habitat LULC parcels impose “edge effects” on habitat parcels and can have negative impacts within habitat parcels by, for example, facilitating entry of predators, competitors, invasive species, or toxic chemicals and other pollutants. Another example: in many developing countries roads are a threat to forest habitat quality on the landscape because of the access they provide to timber and non-timber forest harvesters.

Each threat source needs to be mapped on a raster grid. A grid cell value on a threat’s map can either indicate intensity of the threat within the cell (e.g., road length in a grid cell or cultivated area in a gird cell) or simply a 1 if the grid cell contains the threat in a road or crop field cover and 0 otherwise. Let o_{ry} indicate threat r’s “score” in grid cell y where r = 1, 2, ..., R indexes all modeled degradation sources.

All mapped threats should be measured in the same scale and metric. For example, if one threat is measured in density per grid cell then all degradation sources should be measured in density per grid cell where density is measured with the same metric unit (e.g., km and km2). Or if one threat is measured with presence/absence (1/0) on its map then all threats should be mapped with the presence/absence scale.

The impact of threats on habitat in a grid cell is mediated by four factors.

1. The first factor is the relative impact of each threat. Some threats may be more damaging to habitat, all else equal, and a relative impact score accounts for this (see Table 1 for a list of possible threats). For instance, urban areas may be considered to be twice as degrading to any nearby habitats as agricultural areas. A degradation source’s weight, w_r, indicates the relative destructiveness of a degradation source to all habitats. The weight w_r can take on any value from 0 to 1. For example, if urban area has a threat weight of 1 and the threat weight of roads is set equal to 0.5 then the urban area causes twice the disturbance, all else equal, to all habitat types. To reiterate, if we have assigned species group-specific habitat suitability scores to each LULC then the threats and their weights should be specific to the modeled species group.

2. The second mitigating factor is the distance between habitat and the threat source and the impact of the threat across space. In general, the impact of a threat on habitat decreases as distance from the degradation source increases, so that grid cells that are more proximate to threats will experience higher impacts. For example, assume a grid cell is 2 km from the edge of an urban area and 0.5 km from a highway. The impact of these two threat sources on habitat in the grid cell will partly depend on how quickly they decrease, or decay, over space. The user can choose either a linear or exponential distance-decay function to describe how a threat decays over space. The impact of threat r that originates in grid cell y, r_y, on habitat in grid cell x is given by i_{rxy} and is represented by the following equations,

$$i_{rxy} = 1 - \left(\frac{d_{xy}}{d_{r\ \max}}\right) \text{ if linear} \tag{4.1}$$

$$i_{rxy} = exp\left(-\left(\frac{2.99}{d_{r\ \max}}\right) d_{xy}\right) \text{if exponential} \tag{4.2}$$

where d_{xy} is the linear distance between grid cells x and y and $d_{r\ \max}$ is the maximum effective distance of threat r's reach across space. Figure 1 illustrates the relationship between the distance-decay rate for a threat based on the maximum effective distance of the threat (linear and exponential). For example, if the user selects an exponential decline and the maximum impact distance of a threat is set at 1 km, the impact of the threat on a grid cell’s habitat will decline by ~ 50% when the grid cell is 200 m from r’s source. If $i_{rxy} > 0$ then grid cell x is in degradation source ry’s disturbance zone. (If the exponential function is used to describe the impact of degradation source r on the landscape then the model ignores values of i_{rxy} that are very close to 0 in order to expedite the modeling process.) To reiterate, if we have assigned species group-specific habitat suitability scores to each LULC then threat impact over space should be specific to the modeled species group.

3. The third landscape factor that may mitigate the impact of threats on habitat is the level of legal / institutional / social / physical protection from disturbance in each cell. Is the grid cell in a formal protected area? Or is it inaccessible to people due to high elevations? Or is the grid cell open to harvest and other forms of disturbance?

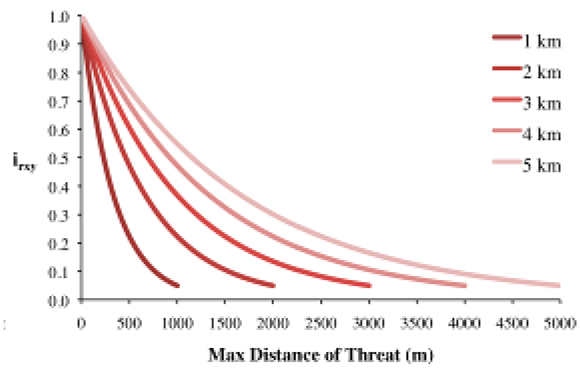


Fig. 4.1: Figure 1. An example of the relationship between the distance-decay rate of a threat and the maximum effective distance of a threat.

The model assumes that the more legal / institutional / social / physical protection from degradation a cell has, the less it will be affected by nearby threats, no matter the type of threat. Let $\beta_x \in [0, 1]$ indicate the level of accessibility in grid cell x where 1 indicates complete accessibility. As decreases the impact that all threats will have in grid cell x decreases linearly. It is important to note that while legal / institutional / social / physical protections often do diminish the impact of extractive activities in habitat such as hunting or fishing, it is unlikely to protect against other sources of degradation such as air or water pollution, habitat fragmentation, or edge effects. If the threats considered are not mitigated by legal / institutional / social / physical properties then you should ignore this input or set $\beta_x = 1$ for all grid cells x. To reiterate, if we have assigned species group-specific habitat suitability scores to each LULC then the threats mitigation weights should be specific to the modeled species group.

4. The relative sensitivity of each habitat type to each threat on the landscape is the final factor used when generating the total degradation in a cell with habitat (in Kareiva et al. 2010 habitat sensitivity is referred to by its inverse, "resistance"). Let $S_{jr} \in [0, 1]$ indicate the sensitivity of LULC (habitat type) j to threat r where values closer to 1 indicate greater sensitivity. The model assumes that the more sensitive a habitat type is to a threat, the more degraded the habitat type will be by that threat. A habitat's sensitivity to threats should be based on general principles from landscape ecology for conserving biodiversity (e.g., Forman 1995; Noss 1997; Lindenmayer et al 2008). To reiterate, if we have assigned species group-specific habitat suitability scores to each LULC then habitat sensitivity to threats should be specific to the modeled species group.

Therefore, the total threat level in grid cell x with LULC or habitat type j is given by D_{xj},

$$D_{xj} = \sum_{r=1}^{R} \sum_{y=1}^{Y_r} \left(\frac{w_r}{\sum_{r=1}^{R} w_r} \right) r_y i_{rxy} \beta_x S_{jr} \tag{4.3}$$

where y indexes all grid cells on r's raster map and Y_r indicates the set of grid cells on r's raster map. Note that each threat map can have a unique number of grid cells due to variation in raster resolution If $S_{jr} = 0$ then D_{xj} is not a function of threat r. Also note that threat weights are normalized so that the sum across all threats weights equals 1.

By normalizing weights such that they sum to 1 we can think of D_{xj} as the weighted average of all threat levels in grid cell x. The map of D_{xj} will change as the set of weights we use change. Please note that two sets of weights will only differ if the relative differences between the weights in each set differ. For example, set of weights of 0.1, 0.1, and 0.4 are the same as the set of weights 0.2, 0.2, and 0.8.

A grid cell's degradation score is translated into a habitat quality value using a half saturation function where the user must determine the half-saturation value. As a grid cell's degradation score increases its habitat quality decreases. Let the quality of habitat in parcel x that is in LULC j be given by Q_{xj} where,

$$Q_{xj} = H_j \left(1 - \left(\frac{D_{xj}^z}{D_{xj}^z + k^z} \right) \right) \tag{4.4}$$

Threat	Number of species endangered by threat, as indicated by Lowe et al. (1990), Moseley (1992), and Beacham (1994)	Estimated number of species endangered by threat, derived by extrapolation of 5% sample from *Federal Register*
Interactions with non-native species	305	340
Urbanization	275	340
Agriculture	224	260
Outdoor recreation and tourism development	186	200
Domestic livestock and ranching activities	182	140
Reservoirs and other running water diversions	161	240
Modified fire regimes and silviculture	144	80
Pollution of water, air, or soil	144	140
Mineral, gas, oil, and geothermal extraction or exploration	140	140
Industrial, institutional, and military activities	131	220
Harvest, Intentional and incidental	120	220
Logging	109	80
Road presence, construction, and maintenance	94	100
Loss of genetic variability, inbreeding depression, or hybridization	92	240
Aquifer depletion, wetland draining or filling	77	40
Native species interactions, plant succession	77	160
Disease	19	20
Vandalism (destruction without harvest)	12	0

Fig. 4.2: Table 1. Possible degradation sources based on the causes of endangerment for American species classified as threatened or endangered by the US Fish and Wildlife Service. Adapted from Czech et al. 2000.

and z (we hard code $z = 2.5$) and k are scaling parameters (or constants). Q_{xj} is equal to 0 if Hj = 0. Q_{xj} increases in Hj and decreases in D_{xj}. Q_{xj} can never be greater than 1. The k constant is the half-saturation constant and is set by the user. The parameter k is equal to the D value where $1 - \left(\frac{D_{xj}^z}{D_{xj}^z + k^z} = 0.5\right)$. For example, if $k = 5$ then $1 - \left(\frac{D_{xj}^z}{D_{xj}^z + k^z}\right) = 0.5$ when $D_{xj} = 5$. In the biodiversity model interface we set $k = 0.5$ but the user can change it (see note in Data Needs section, #8). If you are doing scenario analyses, whatever value you chose for k the first landscape you run the model on, that same k must be used for all alternative scenarios on the same landscape. Similarly, whatever spatial resolution you chose the first time you run the model on a landscape use the same value for all additional model runs on the same landscape. If you want to change your choice of k or the spatial resolution for any model run then you have to change the parameters for all model runs, if you are comparing multiple scenarios on the same landscape.

Habitat Rarity

While mapping habitat quality can help to identify areas where biodiversity is likely to be most intact or imperiled, it is also critical to evaluate the relative rarity of habitats on the landscape regardless of quality. In many conservation plans, habitats that are rarer are given higher priority, simply because options and opportunities for conserving them are limited and if all such habitats are lost, so too are the species and processes associated with them.

The relative rarity of a LULC type on a current or projected landscape is evaluated vis-a-vis a baseline LULC pattern. A rare LULC type on a current or projected map that is also rare on some ideal or reference state on the landscape (the baseline) is not likely to be in critical danger of disappearance, whereas a rare LULC type on a current or projected map that was abundant in the past (baseline) is at risk.

In the first step of the rarity calculation we take the ratio between the current or projected and past (baseline) extents of each LULC type j. Subtracting this ratio from one, the model derives an index that represents the rarity of that LULC class on the landscape of interest.

$$R_j = 1 - \frac{N_j}{N_{j_{\mathrm{baseline}}}} \tag{4.5}$$

where N_j is the number of grid cells of LULC j on the current or projected map and $N_{j_{\mathrm{baseline}}}$ gives the number of grid cells of LULC j on the baseline landscape. The calculation of R_j requires that the baseline, current, and/or projected LULC maps are all in the same resolution. In this scoring system, the closer to 1 a LULC's R score is, the greater the likelihood that the preservation of that LULC type on the current or future landscape is important to biodiversity conservation. If LULC j did not appear on the baseline landscape then we set $R_j = 0$.

Once we have a R_j measure for each LULC type, we can quantify the overall rarity of habitat type in grid cell x with:

$$R_x = \sum_{x=1}^{X} \sigma_{xj} R_j \tag{4.6}$$

where $\sigma_{xj} = 1$ if grid cell x is in LULC j on a current or projected landscape and equals 0 otherwise.

Limitations and Simplifications

In this model all threats on the landscape are additive, although there is evidence that, in some cases, the collective impact of multiple threats is much greater than the sum of individual threat levels would suggest.

Because the chosen landscape of interest is typically nested within a larger landscape, it is important to recognize that a landscape has an artificial boundary where the habitat threats immediately outside of the study boundary have been clipped and ignored. Consequently, threat intensity will always be less on the edges of a given landscape. There are two ways to avoid this problem. One, you can choose a landscape for modeling purposes whose spatial extent is significantly beyond the boundaries of your landscape of interest. Then, after results have been generated, you can extract the results just for the interior landscape of interest. Or the user can limit themselves to landscapes where degradation sources are concentrated in the middle of the landscape.

4.1.4 Data Needs

The model uses seven types of input data (five are required).

1. **Current LULC map (required).** A GIS raster dataset, with a numeric LULC code for each cell. The LULC raster should include the area of interest, as well as a buffer of the width of the greatest maximum threat distance. Otherwise, locations near the edge of the area of interest may have inflated habitat quality scores, because threats outside the area of interested are not properly accounted for. The dataset should be in a projection where the units are in meters and the projection used should be defined.

   Name: it can be named anything.

   Format: standard GIS raster file (e.g., ESRI GRID or IMG), with LULC class code for each cell (e.g., 1 for forest, 2 for agriculture, 3 for grassland, etc.). The LULC class codes should be in the grid's 'value' column. The raster should not contain any other data. The LULC codes must match the codes in the "Sensitivity of land cover types to each threat" table below (input # 7).

   Sample Data Set: \InVEST\HabitatQuality\Input\lc_samp_cur_b

2. **Future LULC map (optional):** A GIS raster dataset that represents a future projection of LULC in the landscape. This file should be formatted exactly like the "current LULC map" (input #1). LULC that appears on the current and future maps should have the same LULC code. LULC types unique to the future map should have codes not used in the current LULC map. Again, the LULC raster should include the area of interest, as well as a buffer of the width of the greatest maximum threat distance. Otherwise, locations near the edge of the area of interest may have inflated habitat quality scores, because threats outside the area of interested are not properly accounted for.

   Name: it can be named anything.

   Format: standard GIS raster file (e.g., ESRI GRID or IMG), with LULC class code for each cell (e.g., 1 for forest, 3 for grassland, etc.). The LULC class codes should be in the raster's 'value' column.

   Sample data set: \InVEST\HabitatQuality\Input\lc_samp_fut_b

3. **Baseline LULC map (optional):** A GIS raster dataset of LULC types on some baseline landscape with a numeric LULC code for each cell. This file should be formatted exactly like the "current LULC map" (input #1). The LULCs that are common to the current or future and baseline landscapes should have the same LULC code across all maps. LULC types unique to the baseline map should have codes not used in the current or future LULC map. Again, the LULC raster should include the area of interest, as well as a buffer of the width of the greatest maximum threat distance. Otherwise, locations near the edge of the area of interest may have inflated habitat quality scores, because threats outside the area of interested are not properly accounted for.

If possible the baseline map should refer to a time when intensive mamagement of the land was relatively rare. For example, a map of LULC in 1851 in the Willamette Valley of Oregon, USA, captures the LULC pattern on the landscape before it was severely modified to for massive agricultural production. Granted this landscape had been modified by American Indian land clearing practices such as controlled fires.

   Name: it can be named anything.

   Format: standard GIS raster file (e.g., ESRI GRID or IMG), with LULC class code for each cell (e.g., 1 for forest, 3 for grassland, etc.). The LULC class codes should be in the grid 'value' column.

   Sample data set: \InVEST\HabitatQuality\Input\lc_samp_bse_b

4. **Threat data (required):** A CSV table of all threats you want the model to consider. The table contains information on the each threat's relative importance or weight and its impact across space.

   Name: file can be named anything

   File Type: *.csv

   Rows: each row is a degradation source

Columns: each column contains a different attribute of each degradation source, and must be named as follows:

1. THREAT: the name of the specific threat. **Threat names must not exceed 8 characters.**
2. MAX_DIST: the maximum distance over which each threat affects habitat quality (measured in km). The impact of each degradation source will decline to zero at this maximum distance.
3. WEIGHT: the impact of each threat on habitat quality, relative to other threats. Weights can range from 1 at the highest, to 0 at the lowest.
4. DECAY: the type of decay over space for the threat. Can have the value of either "linear" or "exponential".

Sample Data Set: \Invest\HabitatQuality\Input\threats_samp.csv

Example: Hypothetical study with three threats. Agriculture degrades habitat over a larger distance than roads do, and has a greater overall magnitude of impact. Further, paved roads attract more traffic than dirt roads and thus are more destructive to nearby habitat than dirt roads.

THREAT	MAX_DIST	WEIGHT	DECAY
dirt_rd	2	0.1	linear
Paved_rd	4	0.4	exponential
Agric	8	1	linear

5. **Sources of threats(s) (required):** GIS raster file of the distribution and intensity of each individual threat. You will have as many of these maps as you have threats. These threat maps should cover the area of interest, as well as a buffer of the width of the greatest maximum threat distance. Otherwise, locations near the edge of the area of interest may have inflated habitat quality scores, because threats outside the area of interested are not properly accounted for. Each cell in the raster contains a value that indicates the density or presence of a threat within it (e.g., area of agriculture, length of roads, or simply a 1 if the grid cell is a road or crop field and 0 otherwise). All threats should be measured in the same scale and units (i.e., all measured in density terms or all measured in presence/absence terms and not some combination of metrics). The extent and resolution of these raster datasets does not need to be identical to that of the scenario maps (the LULCs map from inputs #1, #2, or #3). In cases where the threats and LULC map resolutions vary, the model will use the resolution and extent of the LULC cover map. InVEST will not prompt you for these rasters in the tool interface. It will instead automatically find and use each one, based on names in the "Threats data" table (input # 4). Therefore, these threat maps need to be in a file named "input" that is one level below the workspace identified in the model interface (see below).

Please do not leave any area on the threat maps as 'No Data'. If an area has not threat set the area's threat level equal to 0.

If you are analyzing habitat quality for more than one LULC scenario (e.g., a current and future map or a baseline, current, and future map) then you need a set of threat layers for each modeled scenario. Add a "c" at the end of the raster for all "current" threat layers, a "f" for all future threat layers, and a "b" for all "baseline" threat layers. If you do not use such endings then the model assumes the degradation source layers correspond to the current map. If a threat noted in the Threats data table (input # 4) is inappropriate for the LULC scenario that you are analyzing (e.g., industrial development on a Willamette Valley pre-settlement map from 1851) then enter a threat map for that time period that has all 0 values. If you do not include threat maps for a submitted LULC scenario then the model will not calculate habitat quality on the scenario LULC map.

Finally, note that we assume that the relative weights of threats and sensitivity of habitat to threats do not change over time (we only submit one Threat data table and one Habitat sensitivity data table (inputs # 4 and # 7)). If you want to change these over time then you will have to run the model multiple times.

Name: the name of each raster file should exactly match the name of a degradation source in the rows of the Threats data table (input #2) above with the added "_b" (baseline), "_c" (current), or "_f" (future) to indicate the threat map's period. File name cannot be longer than 7 characters if using a GRID format.

Format: standard GIS raster file (e.g., ESRI GRID or IMG), with a relative degradation source value for each cell from that particular degradation source. The "Value" column indicates the relative degradation source that cell shows. File location: files must be saved in a folder titled "input" within the model's workspace (see below).

Sample data sets: \Invest\HabitatQuality\Input\crp_c; crp_f; rr_c; rr_f; urb_c; urb_f; rot_c; rot_f; prds_c; prds_f; srds_c; srds_f; lrds_c; lrds_f. By using these sets of inputs we are running a habitat quality analysis for the current and future LULC scenario maps. A habitat quality map will not be generated for the baseline map because we have not submitted any threat layers for the baseline map. The name 'crp' refers to cropland, 'rr' to rural residential, 'urb' to urban, 'rot' to rotation forestry, 'prds' to primary roads, 'srds' to secondary roads, and 'lrds' to light roads.

6. **Accessibility to sources of degradation (optional):** A GIS polygon shapefile containing data on the relative protection that legal / institutional / social / physical barriers provide against threats. Polygons with minimum accessibility (e.g., strict nature reserves, well protected private lands) are assigned some number less than 1, while polygons with maximum accessibility (e.g., extractive reserves) are assigned a value 1. These polygons can be land management units or a regular array or hexagons or grid squares. Any cells not covered by a polygon will be assumed to be fully accessible and assigned values of 1.

   File type: GIS polygon shapefile.

   Name: file can be named anything.

   Rows: each row is a specific polygon on the landscape

   Columns:

   1. *ID*: unique identifying code for each polygon. FID also works.
   2. *ACCESS*: values between 0 and 1 for each parcel, as described above.

   Sample data set: \InVEST\HabitatQuality\Input\access_samp.shp

7. Habitat types and sensitivity of habitat types to each threat (required). A CSV table of LULC types, whether or not they are considered habitat, and, for LULC types that are habitat, their specific sensitivity to each threat.

   Name: file can be named anything

   File type: *.csv

   Rows: each row is a LULC type.

   Columns: columns contain data on land use types and their sensitivities to threatss. Columns must be named according to the naming conventions below.

   1. *LULC*: numeric code for each LULC type. Values must match the codes used in the LULC maps submitted in inputs # 1 through 3. All LULC types that appear in the current, future, or baseline maps (inputs # 1 through 3) need to appear as a row in this table.
   2. *NAME*: the name of each LULC
   3. *HABITAT*: Each LULC is assigned a habitat score, Hj, from 0 to 1. If you want to simply classify each LULC as habitat or not without reference to any particular species group then use 0s and 1s where a 1 indicates habitat. Otherwise, if sufficient information is available on a species group's habitat preferences, assign LULC a relative habitat suitability score from 0 to 1 where 1 indicates the highest habitat suitability. For example a grassland songbird may prefer a native prairie habitat above all other habitat types (prairie is given a "Habitat" score of 1 for grassland birds), but will also use a managed hayfield or pasture in a pinch (managed hayfield and pasture is given a "Habitat" score of 0.5 for grassland birds).
   4. *L_THREAT1, L_THREAT2*, etc.: The relative sensitivity of each habitat type to each threat. You will have as many columns named like this as you have threat, and the italicized portions of names must match row names in the "Threat data" table noted above (input # 4). Values range from 0 to

1, where 1 represents high sensitivity to a threat and 0 represents no sensitivity. Note: Even if the LULC is not considered habitat, do not leave its sensitivity to each threat as Null or blank, instead enter a 0 and the model will convert it to NoData.

Sample data set: \Invest\HabitatQuality\Input\sensitivity_samp.csv

Example: A hypothetical study with four LULC and three threats. In this example we treat woodlands and forests as (absolute) habitat and bare soil and cultivated areas as (absolute) non-habitat. Forest mosaic is the most sensitive (least resistant) habitat type, and is more sensitive to dirt roads than paved roads or agriculture (0.9 versus 0.5 and 0.8). We enter 0's across all threats for the two developed land covers, base soil and cultivation.

LULC	NAME	HABITAT	L_AG	L_ROAD	L_DIRT_RD
1	Bare Soil	0	0	0	0
2	Closed Woodland	1	0.5	0.2	0.4
3	Cultivation	0	0	0	0
4	Forest Mosaic	1	0.8	0.8	0.5

8. **Half-saturation constant (required):** This is the value of the parameter k in equation (4). By default it is set to 0.5 but can be set equal to any positive number. In general, you want to set k to half of the highest grid cell degradation value on the landscape. To perform this model calibration you will have to the run the model once to find the highest degradation value and set k for your landscape. For example, if a preliminary run of the model generates a degradation map where the highest grid-cell degradation level is 1 then setting k at 0.5 will produce habitat quality maps with the greatest variation on the 0 to 1 scale (this helps with visual representation of heterogeneity in quality across the landscape). It is important to note that the rank order of grid cells on the habitat quality metric is invariant to your choice of k. The choice of k only determines the spread and central tendency of habitat quality scores. Please make sure to use the same value of k for all runs that involve the same landscape. If you want to change your choice of k for any model run then you have to change the parameters for all model runs.

4.1.5 Running the Model

The model is available as a standalone application accessible from the Windows start menu. For Windows 7 or earlier, this can be found under *All Programs -> InVEST +VERSION+ -> Habitat Quality*. Windows 8 users can find the application by pressing the windows start key and typing "habitat quality" to refine the list of applications. The standalone can also be found directly in the InVEST install directory under the subdirectory *invest-3_x86/invest_habitat_quality.exe*.

Viewing Output from the Model

Upon successful completion of the model, a file explorer window will open to the output workspace specified in the model run. This directory contains an *output* folder holding files generated by this model. Those files can be viewed in any GIS tool such as ArcGIS, or QGIS. These files are described below in Section *Interpreting Results*.

Interpreting Results

Final results are found in the *Output* folder within the *Workspace* specified for this module.

- **Parameter log**: Each time the model is run, a text (.txt) file will appear in the *Output* folder. The file will list the parameter values for that run and will be named according to the service, the date and time, and the suffix.
- **degrad_cur** – Relative level of habitat degradation on the current landscape. A high score in a grid cell means habitat degradation in the cell is high relative to other cells. Grid cells with non-habitat land cover (LULC with Hj = 0) get a degradation score of 0. This is a mapping of degradation scores calculated with equation (3).

- **qual_cur** – Habitat quality on the current landscape. Higher numbers indicate better habitat quality vis-a-vis the distribution of habitat quality across the rest of the landscape. Areas on the landscape that are not habitat get a quality score of 0. This quality score is unitless and does not refer to any particular biodiversity measure. This is a mapping of habitat qulaity scores calculated with equation (4).
- **rarity_cur** – Relative habitat rarity on the current landscape vis-a-vis the baseline map. This output is only created if a baseline LULC map is submitted (input # 3). This map gives each grid cell's value of Rx (see equation (6)). The rarer the habitat type in a grid cell is vis-a-vis its abundance on the baseline landscape, the higher the grid cell's rarity_cur value.

Optional Output Files

If you are running a future scenario (i.e., you have provided input # 2 and future LULC scenario threat layers), you will also see *degrad_fut* and *qual_fut* in the output folder as well. Further, if you have submitted a baseline LULC map (input # 3) as well, you will also see the raster *rarity_fut* in the output folder.

If you have entered a baseline map (input # 3) and threat layers for the baseline (input # 4)), then you will find the rasters *degrad_bse* AND *qual_bse* in the output folder.

Recall, if you are setting Hj for all LULC j on a continuum between 0 and 1 based on the habitat suitability for a particular species group then these results are only applicable to that species group.

Modifying Output and Creating a Landscape Biodiversity Score

The model output doesn't provide landscape-level quality and rarity scores for comparing the baseline, current, and future LULC scenarios. Instead the user must summarize habitat extent and quality and rarity scores for each landscape. At the simplest level, a habitat quality landscape score for a LULC scenario is simply the aggregate of all grid cell-level scores under the scenario. In other words, we can sum all grid-level quality scores on the *qual_bse* (if available), *qual_cur*, and *qual_fut* (if available) maps and then compare scores. A map may have a higher aggregate quality score for several reasons. For one, it may just have more habitat area. However, if the amount of habitat across any two scenarios is approximately the same then a higher landscape quality score is indicative of better overall quality habitat.

Scores for certain areas on a landscape could also be compared. For example, we could compare aggregate habitat quality scores in areas of the landscape that are known to be in the geographic ranges of species of interest. For example, suppose we have geographic range maps of 9 species and have submitted current and future LULC scenario maps to the habitat quality model. In this case we would determine 18 aggregate habitat quality scores, once for each modeled species under each scenario. Let $G_{s_{\mathrm{cur}}}$ indicate the set of grid cells on the current landscape that are in s‘ range. Then the average habitat quality score in species s‘ range on the current landscape is given by,

$$Q_{s_{\mathrm{cur}}} = \frac{\sum_{x=1}^{G^{s_{\mathrm{cur}}}} Q_{xj_{\mathrm{cur}}}}{G^{s_{\mathrm{cur}}}} \tag{4.7}$$

where $Q_{xj_{cur}}$ indicates the habitat quality score on parcel x in LULC j on the current landscape and $Q_{xj_{cur}=0}$ if qual_cur for x is "No Data". The average range-normalized habitat quality score for all 9 species on the current landscape would be given by,

$$R_x = \sum_{x=1}^{X} \sigma_{xj} R_j \tag{4.8}$$

Then we would repeat for the future landscape with the grid cells in set Gs_fut for each species s and the set of $Q_{xj_{fut}}$.

4.1.6 References

Ando, A, J. Camm, S. Polasky, and A. Solow. 1998. Species distributions, land values, and efficient conservation. Science 279:2126-2128.

Czech, B., P. R. Krausman, and P. K. Devers. 2000. Economic Associations among Causes of Species Endangerment in the United States. Bioscience 50:593-601.

Forman, R. 1995. Land Mosaics: The Ecology of landscapes and regions. Cambridge Univ Press. New York.

Forman, R. 2003. Road ecology: science and solutions. Island Press. New York, New York.

Franklin, J.F. and D. B. Lindenmayer. 2009. Importance of matrix habitats in maintaining biological diversity. Proceedings of the National Academy of Sciences 106:349-350.

Hall, L.S., Krausman, P.R. and Morrison, M.L. 1997. The habitat concept and a plea for standard terminology. Wildlife Society Bulletin 25(1):173-182.

Lindenmayer, D., Hobbs, R., Montague-Drake, R., Alexandra, J., Bennett, A., Burgman, M., Cae, P., Calhoun, A., Cramer, V., Cullen, P. 2008. A checklist for ecological management of landscapes for conservation. Ecology Letters 11:78-91.

MacArthur, R., E. 0. Wilson. 1967. The theory of island biogeography. Princeton University Press, Princeton, NJ.

Mckinney, M.L. 2002. Urbanization, biodiversity, and conservation. BioScience 52:883-890.

Nelleman C, Kullered L, Vistnes I, Forbes B, Foresman T, Husby E, Kofinas G, Kaltenborn B, Rouaud J, Magomedova M, Bobiwash R, Lambrechts C, Schei P, Tveitdal S, Gron O, Larsen T. 2001. GLOBIO. Global methodology for mapping human impacts on the biosphere. UNEP/DEWA/TR.01-3.

Nelson, E., S. Polasky, D. J. Lewis, A. J. Plantinga, E. Lonsdorf, D. White, D. Bael & J. J. Lawler. 2008. Efficiency of incentives to jointly increase carbon sequestration and species conservation on a landscape. Proc. Nat. Acad. Sci. 105: 9471-9476.

Noss, R. F., M. A. Connell, and D. D. Murphy. 1997. The science of conservation planning: habitat conservation under the endangered species act. Island Press. Prugh, L., K. Hodges, A. Sinclair, and J. Brashares. 2008. Effect of habitat area and isolation on fragmented animal populations. Proceedings of the National Academy of Sciences 105:20770.

Ricketts, T. H. 2001. The Matrix Matters: Effective Isolation in Fragmented Landscapes. American Naturalist 158:87-99.

Vitousek, P. M., H. A. Mooney, J. Lubchenco, and J. M. Melillo. 1997. Human Domination of Earth's Ecosystems. Science 277:494.

Wilcove, D. S., D. Rothstein, J. Dubow, A. Phillips, and E. Losos. 1998. Quantifying Threats to Imperiled Species in the United States. Bioscience 48:607-615.

4.2 Habitat Risk Assessment

4.2.1 Summary

The condition of a habitat is a key determinant of the ecosystem services it can provide. For example, multiple stressors including fishing, climate change, pollution and coastal development threaten the ability of coastal ecosystems to provide the valuable goods and services that people want and need. As human activities continue to intensify, so too does the need for quick, clear and repeatable ways of assessing the risks posed by human activities under various management plans. The InVEST habitat risk assessment (HRA) model allows users to assess the risk posed to coastal and marine habitats by human activities and the potential consequences of exposure for the delivery of ecosystem services and biodiversity. The InVEST HRA model is similar to the InVEST biodiversity model in that both models allow users to identify regions on a landscape or seascape where human impacts are highest. While the biodiversity model is intended to be used to assess how human activities impact biodiversity, the HRA model is better suited to screening the risk of current and future human activities to prioritize management strategies that best mitigate risk. We built and tested the HRA model in marine and coastal systems, and discuss it accordingly, but it easily can be applied to terrestrial systems, or mobile species.

Risk of human activities (e.g., salmon aquaculture, coastal development, etc.) to habitats (e.g., seagrasses, kelp forests, mangroves, reefs) is a function of the exposure of each habitat to each activity and the consequences for each habitat. Exposure to stressors can arise through direct overlap in space and time or through indirect effects (i.e. finfish farms in an enclosed bay may degrade water quality and thus impede eelgrass growth throughout the bay, even if the netpens are not situated directly over eelgrass beds). Consequence depends on the effects of activities on habitat area and density, and the ability of habitats to recover from these effects (i.e. through processes such as recruitment and regeneration). Outputs from the model are useful for understanding the relative risk of human activities and climate change to habitats within a study region and among alternative future scenarios. Model outputs can help identify areas on the seascape where human activities may create trade-offs among ecosystem services by posing risk high enough to compromise habitat structure and function. The model can help to prioritize areas for conservation and inform the design and configuration of spatial plans for both marine and terrestrial systems.

4.2.2 Introduction

Nearshore habitats such as kelp forests and eelgrass meadows provide valuable ecosystem services including the protection of shorelines from storms, nursery habitat for fisheries and carbon storage and sequestration. As these habitats become degraded by human activities, the ecosystem services they provide are threatened. The impacts of human activities in coastal areas, both on land and in the sea, are pervasive in coastal ecosystems. Recent global analyses have revealed that almost no area of the world's oceans is untouched by human impacts (Halpern et al. 2008). Thus, an understanding of the location and intensity of human impacts on nearshore ecosystems is an essential component of informed and successful coastal and ocean management. The InVEST HRA model allows users to assess the threat of human activities to the health of these ecosystems.

InVEST Biodiversity Model vs. InVEST Habitat Risk Assessment Model

A primary goal of conservation is the protection of biodiversity; biodiversity is intricately linked to the production of ecosystem services. While some consider biodiversity itself to be an ecosystem service, the InVEST biodiversity model treats it as an independent attribute of natural systems, with its own intrinsic value (InVEST does not monetize biodiversity). InVEST includes a biodiversity model because natural resource managers, corporations and conservation organizations are becoming increasingly interested in understanding how and where biodiversity and ecosystem services align in space and how management actions affect both. The biodiversity model uses habitat quality and rarity as a proxy for diversity.

When developing a similar model with marine systems in mind, differences in data availability (e.g., the lack of an analog to land-use/land-cover maps in marine systems) and differences in thinking (e.g., the prevalence of a risk-assessment framework in fisheries science) led us to the development of the habitat risk assessment model described in this chapter.

Both the biodiversity model and the HRA model can be used to identify areas on a landscape or seascape where the risk posed by human activities is highest. Indeed, while the two models are segregated into the marine and terrestrial toolboxes, they can be used across systems. However, the modeling approaches differ in several ways. First, the exposure-consequence framework of the HRA model allows model results to be interpreted in a manner that helps users explore which types of management strategies are likely to most effectively reduce risk (Figure 1). For example, ecosystems with high exposure and high consequence may be targeted for intense active management, while effective strategies for ecosystems with low exposure to human stressors but high consequence may include close monitoring but little active intervention unless exposure increases. Second, the transparent flexible structure and explicit visualization of data uncertainty in the HRA model facilitate its use in both data-rich and data-poor situations.

Finally, the biodiversity model is better suited for terrestrial applications than marine applications because it requires a land use land cover map as an input. The HRA model can be used in both marine and terrestrial systems.

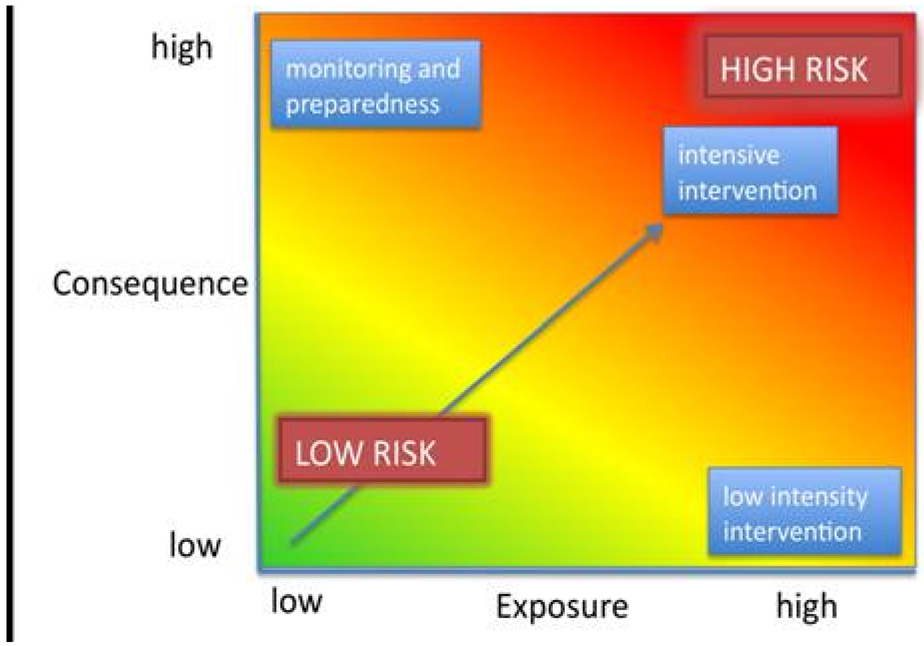


Fig. 4.3: Habitats with high exposure to human activities and high consequence are at high risk. Plotting exposure and consequence data in this plot allows users to visualize risk, and to assess which types of risk are more effectively mitigated by human intervention (risks driven by exogenous human factors, top right region of the risk space) and which types of risk are better addressed through monitoring and preparedness (risks driven by endogenous habitat-specific factors). (Adapted from Dawson et al. 2011).

Risk assessment has a long history in the field of ecotoxicology, and is now emerging as a valuable method in ecosystem-based fisheries management (Astles et al. 2006, Hobday et al. 2011). Risk assessment is used to determine the likelihood that a hazard will cause undesired consequences (Burgman 2005). In the context of marine ecosystem-based management, risk assessment evaluates the probability that human activities will impede the achievement of desired marine management objectives. In the HRA model, we define risk as the likelihood that human activities will reduce the quality of nearshore habitats to the point where their ability to deliver ecosystem services is impaired. Researchers have made significant progress in evaluating human impacts on marine ecosystems in recent years. However many of these approaches lack generality because they are focused on the effects of a single sector (i.e. fisheries e.g. Astles et al. 2006, Hobday et al. 2011), or have limited transparency and flexibility because they are based on expert opinion (Halpern et al. 2008, Teck et al. 2010). The HRA model in Marine InVEST builds on these approaches and allows users to evaluate the risk posed by a variety of human activities to key coastal habitats in a transparent, repeatable and flexible way.

4.2.3 The Model

The risk of human activities to coastal and nearshore habitats is a function of the habitat's exposure to the activity and the consequence of exposure. To determine exposure, users provide model inputs such as base maps of habitat distribution and human activities, the timing and intensity of the activity and the effectiveness of current management practices in safeguarding habitats. To determine consequence, users provide model inputs such as observed loss of habitat and the ability of habitats to recover. The model is flexible and can accommodate data-poor and data-rich situations. Data may come from a combination of peer-reviewed sources at the global scale and locally available fine-scale data sources. Model inputs and results can be updated as better information becomes available.

The HRA model produces information about risk at two scales and with several types of outputs. Maps display variation at a grid cell scale in the relative risk of human activities to habitats within the study area and among alternative future scenarios. Tables and risk plots (ie., Figure 1) show the contribution of different activities to the risk posed to each habitat at a subregional scale within the study area and among future scenarios. When run as part of a complete Marine InVEST analysis, the HRA model can be used to identify which human activities are likely to cause trade-offs in other ecosystem services. As a result, the model will help managers prioritize and evaluate management strategies with regards to their effectiveness of reducing risks to nearshore habitats and maintaining the delivery of desired ecosystem services.

How it Works

The HRA model combines information about the exposure of habitats to each stressor with information about the consequence of that exposure for each habitat to estimate and produce maps of risk to habitats and habitat quality at both a grid cell and a subregional scale. For example, exposure depends on the extent of geographic overlap between habitats and human activities, the duration of time that the activity and habitat overlap, the intensity of the stressor and the degree to which management strategies mitigate impact. Consequence depends on the degree of habitat loss as a result of exposure to a stressor, change in habitat structure and the ability of habitats to recover from these effects (i.e., through life history traits such as recruitment and regeneration rates). The modelling approach is flexible so if any of the default factors that influence exposure and consequence are irrelevant in a particular case, they can be excluded and/or replaced with alternative criteria (see *Risk of Human Activities to Habitats*). We begin by explaining the approach at the grid cell scale and later describe differences in the approach for the subregional outputs.

Risk of Human Activities to Habitats

The risk of human activities to habitats is modeled in four steps.

Step 1. The first step involves determining the likelihood of exposure of the habitat to the stressor and the consequence of this exposure. Exposure (E) and consequence (C) are both determined by assigning a rating (typically 1-3, with 0 = no score) to a set of criteria for each attribute. We have provided the user with a set of standard criteria used frequently in the scientific literature, but any criteria may be added or removed. Guidelines for scoring of the default criterion are summarized below, and abbreviated descriptions of scoring on a 1-3 basis are provided in the tables produced from HRA Preprocessor. Note that we treat "spatial overlap" which is one of the exposure criteria, differently from the other default criteria. For each grid cell in the study area, if a stressor and a habitat overlap in space, then spatial overlap = 1 and the model calculates E and C using the information about the other criteria and the equations below. If a stressor and a habitat do not overlap in a particular grid cell, we assume that spatial overlap = 0, E = 0, C = 0 and Risk = 0. Spatial overlap is determine by the model using the spatial layers for stressor and habitat provided by the user. The scores for all the other criteria are inputs to the model provided by the user. To ensure accuracy, we recommend that scores be determined using readily available data from peer-reviewed literature or published reports, however, you are free to use any data you believe to be the most accurate. For each score assigned, you may also indicate the quality of the data used to determine the score and the weighted importance of the criteria relative to other criteria. This allows you to assign greater weight to criteria where scoring confidence was higher, or to criteria which contribute more to risk in the system. Thus, the overall exposure E and consequence C scores are calculated as weighted averages of the exposure values e_i and consequence values c_i for each criterion *i* as

$$E = \frac{\sum_{i=1}^{N} \frac{e_i}{d_i \cdot w_i}}{\sum_{i=1}^{N} \frac{1}{d_i \cdot w_i}} \tag{4.9}$$

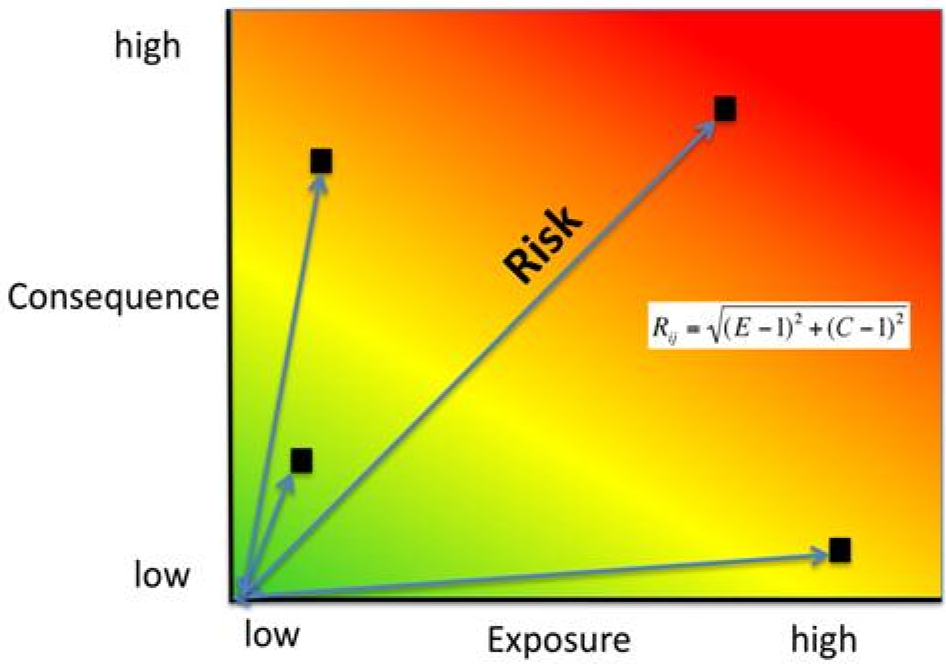
$$C = \frac{\sum_{i=1}^{N} \frac{c_i}{d_i \cdot w_i}}{\sum_{i=1}^{N} \frac{1}{d_i \cdot w_i}} \tag{4.10}$$

where d_i represents the data quality rating for criterion *i*, w_i represents the importance weighing for criterion *i* and N is the number of criteria evaluated for each habitat.

Step 2. The second step combines the exposure and response values to produce a risk value for each stressor-habitat combination. There are two options for risk calculation.

For Euclidean Risk calculation, risk to habitat *i* caused by stressor *j* is calculated as the Euclidean distance from the origin in the exposure-consequence space,

$$R_{ij} = \sqrt{(E-1)^2 + (C-1)^2} \tag{4.11}$$



For Multiplicative Risk calculation, risk to habitat *i* caused by stressor *j* is calculated as the product of the summed exposure and consequence scores.

$$R_{ij} = E \cdot C \tag{4.12}$$

The user has the option of choosing which risk function to use. As discussed in the introduction, several recent papers examine risk to marine fisheries, stocks, habitats and ecosystems (Halpern et al. 2008, Teck et al. 2011, Hobday et al. 2011, Williams et al. 2011, Samhouri and Levin 2012). In each case, 2 axes of information are used to calculate total risk. In the InVEST HRA model, we refer to these axes as exposure and consequence. Initial sensitivity testing suggests that overall, the euclidean and multiplicative approaches will agree on the same highest and lowest risk species and habitats; however, there may be differences in the rank order of species at intermediate risk, depending on the values for E and C. The euclidean approach may provide more conservative, higher overall estimates than the multiplicative approach. If E and C values are widely different then the euclidean approach will produce relatively higher risk results. In contrast, the multiplicative approach will tend to produce relatively lower, less conservative values for risk and associate similarity in E and C with higher risk. If your system contains habitats for which there is a very high consequence of risk but low exposure (e.g., coral and shrimp trawling zones that currently avoid coral reefs) and you want to adopt the precautionary principle, then we advise you to choose the Euclidean approach. Either approach will produce useful results. We have included this explanation to provide you with insight into the assumptions of the different functions.

Step 3. In this step, the model quantifies the cumulative risk of all stressors on the habitats. Cumulative risk for habitat *i* is the sum of all risk scores for each habitat,

$$R_i = \sum_{j=1}^{J} R_{ij} \tag{4.13}$$

Step 4. The model identifies areas of habitats that are risk 'hotspots'. These are areas where the influence of human-derived stressors is so great that ecosystem structure and function may be severely compromised. In these areas, there may be trade-offs between human activities and a range of ecosystem services. Thus, users may choose to consider these habitats to be functionally absent in inputs to other InVEST ecosystem service models (see the Interpreting Results section for guidance on how to use risk hotspots to identify trade-offs among human activities under alternative scenarios). Each grid cell for each type of habitat is classified as HIGH, MED or LOW risk based on risk posed by any individual stressor or the risk posed by the cumulative effects of multiple stressors. A classification of HIGH is assigned to grid cells with a cumulative risk of >66% of the maximum risk score for any individual stressor-habitat combination, or >66% of total possible cumulative risk. Total possible cumulative risk is based on the maximum number of stressors that can occupy a particular grid cell in the study area (see next paragraph). For example, maximum overlapping stressors = 3 if in the entire study region no more than 3 stressors (e.g., agriculture run-off, marine aquaculture and marine transportation) are likely to occur in a single grid cell. Cells are classified as MED if they have individual stressor or cumulative risk scores between 33%-66% of the total possible cumulative risk score. Cells are classified as LOW risk if they have individual or cumulative risk scores of 0-33% of the total possible risk score for a single stressor or multiple stressors, respectively.

The maximum number of overlapping stressor is an input provided by the user. The highest value for this input is the total number of stressors in the study area; however, it is unlikely that all stressors will ever realistically overlap in a single grid cell. This is because stressors are distributed differently in space (i.e., stressors like coastal development exist along the shore while shipping lanes exist offshore) and because some stressors can never exist in the exact same location (i.e., naval weapons testing areas and tourism). From applying this model in several locations, we have found that 3 or 4 is a common value for maximum number of overlapping stressors, but the user should either manually examine his/her maps for overlaps in stressors or use the overlap analysis model to calculate the highest number of overlapping stressors.

Step 5. In the final step, the user has the option of assessing risk at a subregional scale, which is larger than the resolution of the grid cells and smaller than the size of the study area. In a coastal and marine spatial planning process, subregions are often units of governance (i.e., coastal planning regions, states or provinces) within the boundaries of the planning area. Risk outputs at a subregional scale can be used to determine which activities are contributing the most to habitat risk in a particular region. This information can in turn be used to explore strategies that would reduce the exposure of a particular habitat to a particular activity, such as reducing the extent or changing the location of an activity. The model produces risk plots for each habitat that compare the consequence and exposure scores for all activities at a subregional scale. These plots help to user to understand if reducing exposure of particular activities through management actions is likely to reduce risk or if risk is driven by consequence, which is harder to perturb through management actions (see Figure 1 above). The model also produces tables listing E, C and Risk for each habitat-stressor combination at a subregional scale and calculates the percentage of cumulative risk by habitat that is due to a particular stressor in that region. Note that the subregional score for spatial overlap (a default exposure criteria) is based on the fraction of habitat area in a subregion that overlaps with a human activity (see below for more detail). The subregional score for all other E and C criteria are the average E and C score across all grid cells in the study area. Risk is estimated either using the Euclidean distance or multiplicative approach (see above).

Exposure and Consequence Criteria in More Detail

The model allows for any number of criteria to be used when evaluating the risk to habitat areas. As a default, the model provides a set of typical considerations for evaluating risk of stressors to habitats. With the exception of spatial overlap at a grid cell scale, these criteria are rated on a scale of 1-3, with 0 = no score. But we do not constrain the rating of these criteria to a 0-3 scale. If there is significant literature using an alternative scale, the model can accomodate any scale (i.e., 1-5, 1-10) as long as there is consistency across the rating scores within a single model run. It should be noted that using a score of 0 on ANY scale will indicate that the given criteria is not desired within that model run.

Exposure of Habitats to Stressors The risk of a habitat being affected by a stressor depends in part on the exposure of the habitat to that stressor. Stressors may impact habitats directly and indirectly. Because indirect impacts are poorly understood and difficult to trace, we only model the risk of stressors that directly impact habitat by overlapping in space. Other important considerations include the duration of spatial overlap, intensity of the stressor, and whether management strategies reduce or enhance exposure.

1. **Spatial overlap .** To assess spatial overlap in the study area, the model uses maps of the distribution of habitats and stressors. Habitat types can be biotic, such as eelgrass or kelp, or abiotic, such as hard or soft bottom. The user defines the detail of habitat classification. For example, habitats can be defined as biotic or abiotic, by taxa (e.g., coral, seagrass, mangrove), by species (e.g., red, black mangroves) or in whatever scheme the user desires. However, the user should keep in mind that in order for such detail to be useful and change the outcome of the model, these habitat classifications should correspond with differences between habitats in their response to the stressors.

   The model also requires the user to input maps of the distribution of each stressor and information about its "zone of influence." The zone of influence of each stressor is the distance over which the effects of the stressor spread beyond its actual footprint in the input stressor map. For some stressors, such as over-water structures that shade eelgrass beds, this distance will be small. For other stressors, such as finfish aquaculture pens where nutrients spread 300-500m, this distance may be large. The model uses the distance of influence of a stressor to create an intermediate output that is a map of the stressor footprint buffered by the zone of influence. The model uses the maps of habitat and buffered stressors to estimate spatial overlap between each habitat and each stressor at the grid cell and subregional scale.

   For each grid cell, if the habitat overlaps with a stressor, then spatial overlap = 1 and the model calculates exposure, consequence and risk using scores for the other criteria (below). If a habitat does not overlap with a stressor in a particular grid cell, then the model sets exposure, consequence and risk = 0 in that particular grid cell. At the subregional scale, the model calculates the fraction of area of each habitat that overlaps with each stressor. Next the model puts that fraction on a scale of 1- maximum risk score to match the scale for scoring the other criteria. For example, if spatial overlap = 50% of the habitat overlapped by a stressor, and our scale is 0-3, then 3*overlap + 1*(1-overlap) = 2. Lastly, the model averages the spatial overlap score with the average exposure score for the subregion. If there is no spatial overlap between the habitat and stressor at the subregional scale, then exposure = 0, consequence = 0 and risk = 0. If there are no exposure scores for that habitat-stressor combination, but spatial overlap does exist, the score will be entirely the spatial overlap.

2. **Overlap time rating.** Temporal overlap is the duration of time that the habitat and the stressor experience spatial overlap. Some stressors, such as permanent overwater structures, are present year-round; others are seasonal, such as certain fishing practices. Similarly, some habitats (e.g. mangroves) are present year round, while others are more ephemeral (e.g. some seagrasses).

   We use the following categories to classify HIGH, MEDIUM and LOW temporal overlap:

	High (3)	Medium (2)	Low (1)	No score (0)
Temporal overlap	Habitat and stressor co-occur for 8-12 months of the year	Habitat and stressor co-occur for 4-8 months of the year	Habitat and stressor co-occur for 0-4 months of the year	N/A

   Choose "No score" to exclude this criteria from your assessment.

3. **Intensity rating.** The exposure of a habitat to a stressor depends not only on whether the habitat and stressor overlap in space and time, but also on the intensity of the stressor. The intensity criterion is stressor-specific. For example, the intensity of nutrient-loading stress associated with netpen salmon aquaculture is related to the number of salmon in the farm and how much waste is released into the surrounding environment. Alternatively, the intensity of destructive shellfish harvesting is related to the number of harvesters and the harvest practices. You can use this intensity criteria to explore how changes in the intensity of one stressor might affect risk to habitats. For example, one could change the intensity score to represent changes in the stocking density of a salmon farm in a future scenario. One can also use this ranking to incorporate relative differences in the intensity

of different stressors within the study region. For example, different types of marine transportation may have different levels of intensity. For example, cruise ships may be a more intense stressor than water taxis because they release more pollutants than the taxis do.

We use the following categories to classify HIGH, MEDIUM and LOW intensity:

	High (3)	Medium (2)	Low (1)	No score (0)
Intensity	High intensity	Medium intensity	Low intensity	N/A

Choose "No score" to exclude this criteria from your assessment.

4. **Management strategy effectiveness rating.** Management can limit the negative impacts of human activities on habitats. For example, policies that require salmon aquaculturists to let their farms lie fallow may reduce the amount of waste released and allow nearby seagrasses to recover. Similarly, regulations that require a minimum height for overwater structures reduce the shading impacts of overwater structures on submerged aquatic vegetation. Thus, effective management strategies will reduce the exposure of habitats to stressors. The effectiveness of management of each stressor is scored relative to other stressors in the region. So if there is a stressor that is very well managed such that it imparts much less stress on the system than other stressors, classify management effectiveness as "very effective." In general, however, the management of most stressors is likely to be "not effective." After all, you are including them as stressors because they are having some impact on habitats. You can then use this criterion to explore changes in management between scenarios, such as the effect of changing coastal development from high impact (which might receive a score of "not effective") to low impact (which might receive a score of "somewhat effective)."

   We use the following categories to classify HIGH, MEDIUM and LOW management effectiveness:

	High (3)	Medium (2)	Low (1)	No score (0)
Management effectiveness	Not effective, poorly managed	Somewhat effective	Very effective	N/A

Choose "No score" to exclude this criteria from your assessment.

Consequence of Exposure The risk of a habitat being degraded by a stressor depends on the consequence of exposure. The consequence of exposure depends on the ability of a habitat to resist the stressor and recover following exposure, and can be assessed using four key attributes: change in area, change in structure, frequency of natural disturbance, and recovery attributes. We describe each in turn below.

1. **Change in area rating.** Change in area is measured as the percent change in areal extent of a habitat when exposed to a given stressor and thus reflects the sensitivity of the habitat to the stressor. Habitats that lose a high percentage of their areal extent when exposed to a given stressor are highly sensitive, while those habitats that lose little area are less sensitive and more resistant.

   We use the following categories to classify HIGH, MEDIUM and LOW change in area:

	High (3)	Medium (2)	Low (1)	No score (0)
Change in area	High loss in area (50-100%)	Medium loss in area (20-50%)	Low loss in area (0-20%)	N/A

Choose "No score" to exclude this criteria from your assessment.

2. **Change in structure rating.** For biotic habitats, the change in structure is the percentage change in structural density of the habitat when exposed to a given stressor. For example, change in structure would be the change in shoot density for seagrass systems, change in polyp density for corals, or change in stipe density for kelp systems. Habitats that lose a high percentage of their structure when exposed to a given stressor are highly sensitive, while habitats that lose little structure are less sensitive and more resistant. For abiotic habitats, the change in structure is the amount of structural damage sustained by the habitat. Sensitive abiotic habitats will sustain complete or partial damage, while those that sustain little to no damage are more resistant. For example,

gravel or muddy bottoms will sustain partial or complete damage from bottom trawling while hard bedrock bottoms will sustain little to no damage.

We use the following categories to classify HIGH, MEDIUM and LOW change in structure:

	High (3)	Medium (2)	Low (1)	No score (0)
Change in structure	High loss in structure (for biotic habitats, 50-100% loss in density, for abiotic habitats, total structural damage)	Medium loss in structure (for biotic habitats, 20-50% loss in density, for abiotic habitats, partial structural damage)	Low loss in structure (for biotic habitats, 0-20% loss in density, for abiotic habitats, little to no structural damage)	N/A

Choose "No score" to exclude this criteria from your assessment.

3. **Frequency of natural disturbance rating.** If a habitat is naturally frequently perturbed in a way similar to the anthropogenic stressor, it may be more resistant to additional anthropogenic stress. For example, habitats in areas that experience periodical delivery of nutrient subsidies (i.e. from upwelling or allocthonous inputs such as delivery of intertidal plant material to subtidal communities) are adapted to variable nutrient conditions and may be more resistant to nutrient loading from netpen salmon aquaculture. This criterion is scored separately for each habitat-stressor combination, such that being adapted to variable nutrient conditions increases resistance to nutrient loading from salmon aquaculture but not destructive fishing. However, high storm frequency may increase resistance to destructive fishing, because both stressors impact habitats in similar ways.

   We use the following categories to classify HIGH, MEDIUM and LOW natural disturbance frequencies:

	High (3)	Medium (2)	Low (1)	No score (0)
Frequency of natural disturbance	Annually or less often	Several times per year	Daily to weekly	N/A

   Choose "No score" to exclude this criteria from your assessment.

Note: The following consequence criteria are Recovery Attributes. These include life history traits such as regeneration rates and recruitment patterns influence the ability of habitats to recover from disturbance. For biotic habitats, we treat recovery as a function of natural mortality, recruitment, age of maturity, and connectivity.

4. **Natural mortality rate rating (biotic habitats only).** Habitats with high natural mortality rates are generally more productive and more capable of recovery.

   We use the following categories to classify HIGH, MEDIUM and LOW natural mortality rates:

	High (3)	Medium (2)	Low (1)	No score (0)
Natural mortality rate	Low mortality (e.g. 0-20%)	Moderate mortality (e.g. 20-50%)	High mortality (e.g.80% or higher)	N/A

   Choose "No score" to exclude this criteria from your assessment.

5. **Recruitment rating (biotic habitats only).** Frequent recruitment increases recovery potential by increasing the chance that incoming propagules can re-establish a population in a disturbed area.

   We use the following categories to classify HIGH, MEDIUM and LOW natural recruitment rate:

	High (3)	Medium (2)	Low (1)	No score (0)
Natural recruitment rate	Every 2+ yrs	Every 1-2 yrs	Annual or more often	N/A

   Choose "No score" to exclude this criteria from your assessment.

6. **Age at maturity/recovery time.** Biotic habitats that reach maturity earlier are likely to be able to recover more quickly from disturbance than those that take longer to reach maturity. Here we refer to maturity of the habitat as a whole (i.e., a mature kelp forest) rather than reproductive maturity of individuals. For abiotic habitats, shorter recovery times for habitats such as mudflats decrease the consequences of exposure to human activities. In contrast, habitats made of bedrock will only recover on geological time scales, greatly increasing the consequences of exposure.

   We use the following categories to classify HIGH, MEDIUM and LOW age at maturity/recovery time:

	High (3)	Medium (2)	Low (1)	No score (0)
Age at maturity/recovery time	More than 10 yrs	1-10yrs	Less than 1 yr	N/A

   Choose "No score" to exclude this criteria from your assessment.

7. **Connectivity rating (biotic habitats only).** Larval dispersal and close spacing of habitat patches increases the recovery potential of a habitat by increasing the chance that incoming propagules can re-establish a population in a disturbed area.

   We use the following categories to classify HIGH, MEDIUM and LOW connectivity:

	High (3)	Medium (2)	Low (1)	No score (0)
Connectivity	Low dispersal (less than 10km)	Medium dispersal (10-100km)	High dispersal (>100km)	N/A

   Choose "No score" to exclude this criteria from your assessment.

Using Spatially Explicit Criteria

As an alternative to assigning a single rating to a criteria that is then applied to the whole study region, the model allows for spatially explicit criteria to be used as an input. Spatially explicit criteria ratings can be used for any of the exposure or consequence criteria. For example, the user could differentiate between areas of high and low recruitment for a particular habitat or species within the study area. As another example, the user may have information on spatial variation in a human activity, such as density of fish in aquaculture pens which could influence the intensity rating of this stressor. The spatially explicit criteria are vector layers, where each feature may contain a separate rating for that particular area. (See the *HRA Preprocessor* section for more information how how to prepare and use spatially explicit criteria within a complete model run.)

Guidelines for Scoring Data Quality

Risk assessment is an integrative process, which requires a substantial amount of data on many attributes of human and ecological systems. It is likely that some aspects of the risk assessment will be supported by high quality data and other aspects will be subject to limited data availability and high uncertainty. The user has the option of scoring data quality to put greater weight on the criteria for which confidence is higher in the calculation of risk (eq. 2 and 3). We hope that by including the option to rate data quality in the model, users will be aware of some sources of uncertainty in the risk assessment, and will therefore be cautious when using results derived from low quality data. In addition, the information generated from this rating process can be used to guide research and monitoring effects to improve data quality and availability. We suggest the users first run the model with the same data quality score (e.g., 1) for all the criteria to determine if the overall patterns make sense based just on relationships between the stressors and habitats. Next, if users do not have verified information on the data quality of a given criteria, they should then re-run the model using a 2 or 3 to indicate lower and lowest possible data quality.

For each exposure and consequence score, users can indicate the quality of the data that were used to determine the score on a sliding scale where 1 indicates a limited knowledge of the data quality, and anything above that would be seen as increasingly trustworthy data.

Best data	Adequate data	Limited data
Substantial information is available to support the score and is based on data collected in the study region (or nearby) for the species in question.	Information is based on data collected outside the study region, may be based on related species, may represent moderate or insignificant statistical relationships.	No empirical literature exists to justify scoring for the species but a reasonable inference can be made by the user.

Limitations and Assumptions

Limitations

1. **Results are limited by data quality**: The accuracy of the model results is limited by the availability and quality of input data. Using high quality data such as those from local assessments replicated at several sites within the study region for the species in question within the last ten years will yield more accurate results than using lower quality data that are collected at a distant location with limited spatial or temporal coverage. In most cases, users will need to use data from other geographic locations for some of the stressor-habitat combinations because most of the data on the effects of some stressors have only been collected in a limited number of locations worldwide. To overcome these data limitations, we include a data quality score in the analysis. This score allows users to down-weight criteria for which data quality is low.
2. **Results should be interpreted on a relative scale**: Due to the nature of the scoring process, results can be used to compare the risk of several human activities among several habitats within the study region (which can range in size from small local scales to a global scale), but should not be used to compare risk calculations from separate analyses.
3. **Results do not reflect the effects of past human activities**. The HRA model does not explicitly account for the effects of historical human activities on the current risk. Exposure to human activities in the past may affect the consequence of human activities in the present and future. If users have historical data on the exposure of habitats to human activities (e.g. spatial and temporal extent), and information on how this affects current consequence scores, they may include this information in the analysis for more accurate results.
4. **Results are based on equal weighting of criteria unless the user weights the criteria by importance or data quality**. The model calculates the exposure and consequence scores assuming that the effect of each criterion (i.e. spatial overlap and recruitment pattern) is of equal importance in determining risk. The relative importance of each of the criteria is poorly understood, so we assume equal importance. However, the user has the option to weight the importance of each criterion in determining overall risk.

Assumptions

1. **Often information in the literature about the effect stressors on habitats comes from only a few locations**. If using globally available data or data from other locations, users make the assumption that *ecosystems around the world respond in similar ways to any given stressor* (i.e. eelgrass in the Mediterranean responds to netpen aquaculture in the same way as eelgrass in British Columbia). To avoid making this assumption across the board, users should use local data whenever possible.
2. **Cumulative risk is additive (vs. synergistic or antagonistic)**. The interaction of multiple stressors on marine ecosystems is poorly understood (see Crain et al. 2008 for more information). Interactions may be additive, synergistic or antagonistic. However, our ability to predict the type of interaction that will occur is limited. Due to the absence of reliable information on the conditions that determine additivity, synergism or antagonism, the model assumes additivity because it is the simplest approach. As a result, the model may over- or under-estimate the cumulative risk depending on the set of stressors occurring in the study region.

4.2.4 Data Needs

The model uses an interface to input all required and optional data and a series of Comma Separated Value (CSV) files with which to score all criteria and their data quality. Here we outline the options presented to the user via the interface and the maps and data tables that will be used by the model. First we describe required inputs, followed by a description of optional inputs.

To run the model, three steps are required:

1. Run the HRA Preprocessor Tool
2. Fill out the Ratings CSVs
3. Run the Habitat Risk Assessment model

HRA Preprocessor

Before running the HRA model, it is necessary to concatenate and rate all applicable criteria information. This can be accomplished by running the Preprocessor tool, then editing the resulting CVSs. If you have already run the model, or have the 'habitat_stressor_ratings' directory from a previous HRA Preprocessor run, you may skip this step and proceed to running the Habitat Risk Assessment tool.

To run the tool, run the HRA Preprocessor executable. This will launch a graphical user interface (GUI).

There are several pieces that should be used as inputs to this tool. At any time, you can click the blue question marks to the right of an input for additional guidance.

1. **Workspace Location (required)**. Users are required to specify a workspace folder path. Running HRA Preprocessor creates a folder named 'habitat_stressor_ratings' within this workspace. This Folder will hold all relevant CSVs for criteria rating in the particular model run.

```
Name: Path to a workspace folder.  Avoid spaces.
Sample path: \InVEST\HabitatRiskAssess_3_0\
```

2. **Calculate Risk To Habitat/Species** Here you will select the habitats and/or species that will be inputs for this run of the model. Each of these inputs should point to a directory containing all of the named habitat or species shapefile layers that you wish to include in this model run. The file names are not required to contain an identifying number. Each directory should be independent of the others so as to avoid incorrect repetition in the outputs, and should contain ONLY layers that are desired within this assessment. All layers must be projected in the same projection.:

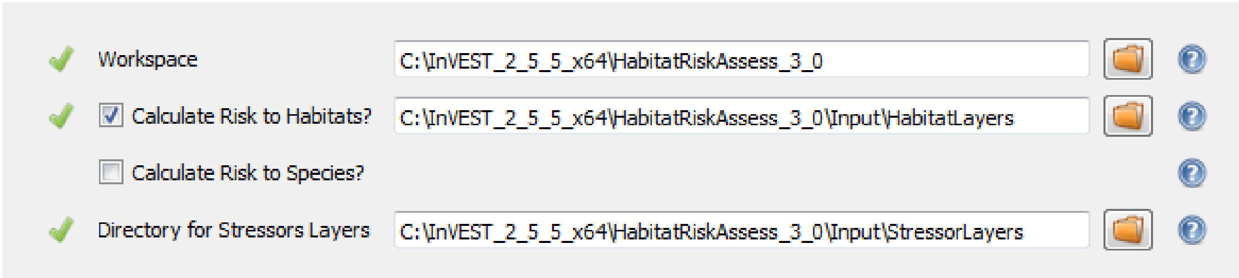
```
Name: Path to a habitat or species folder.  Avoid spaces.
Sample path: \InVEST\HabitatRiskAssess_3_0\HabitatLayers
```

3. **Directory for Stressor Layers** Users should select a folder containing stressors to be overlapped with habitats and/or species. This directory should contain ONLY the stressors desired within this model run. All layers must be projected in the same projection.:

```
Name: Path to a habitat or species folder.  Avoid spaces.
Sample path: \InVEST\HabitatRiskAssess_3_0\StressorLayers
```

4. **Criteria** We have divided up criteria into 3 categories: Exposure, Sensitivity, and Resilience. Exposure criteria are specific to a habitat-stressor pairing, and will be applied to the exposure portion of the risk modeling equation. Sensitivity criteria are also applied to a specific habitat-stressor pairing, but will be applied to the

Fig. 4.4: The HRA Preprocessor main user interface.



consequence portion of the risk equation. Resilience criteria will likewise be applied to the consequence portion of the risk equation, but are specific to an overall habitat. Placing a criteria into one of these categories within the user interface will determine how user ratings are input into the HRA model. The default criteria provided are derived from peer-reviewed literature and are recommended as a good set of contributors to risk in a system, but users do have the option to add or remove criteria if desired. Only choose this option if the default criteria do not apply to the system being modeled, or do not correctly address all facets of the risk assessment.

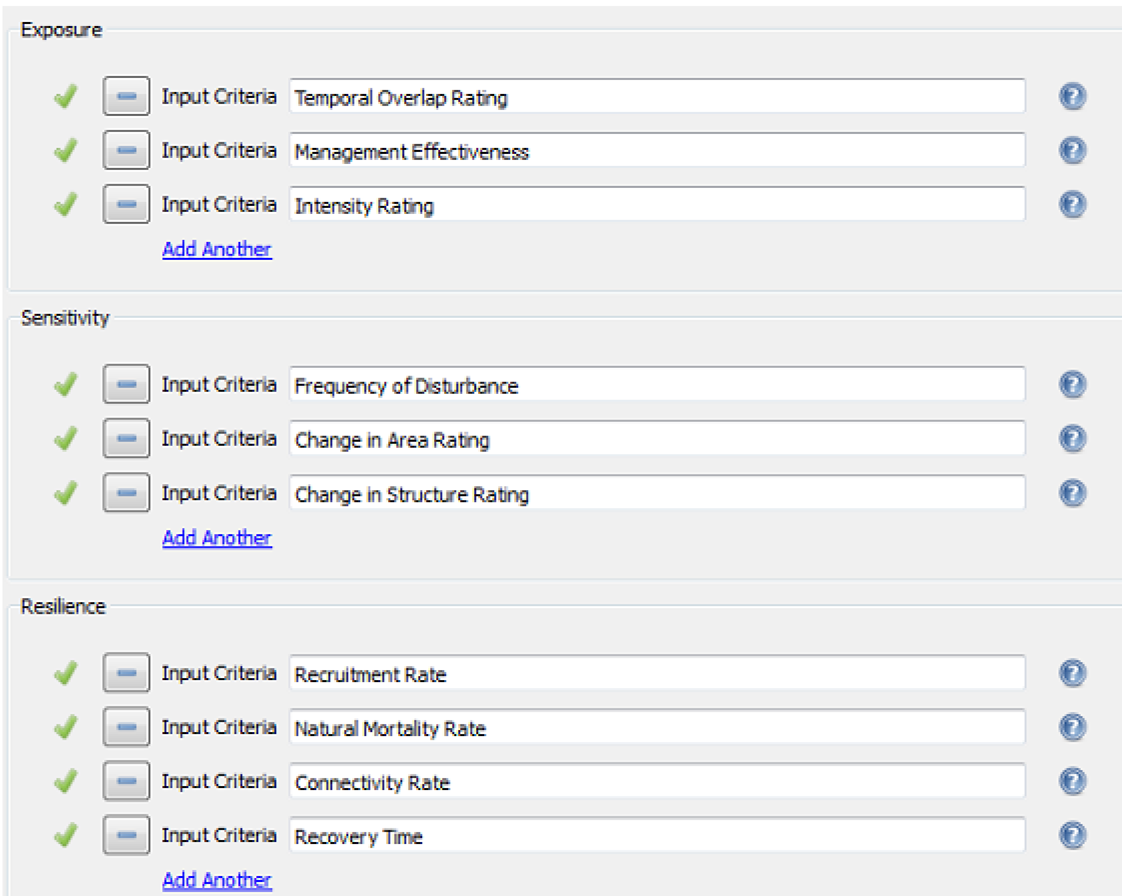
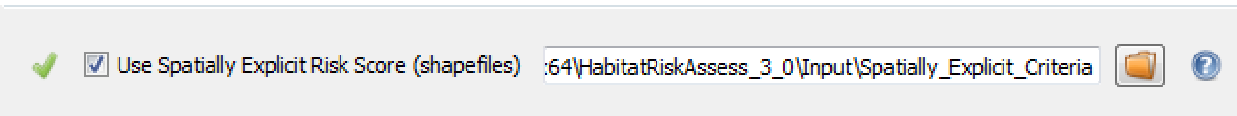


Fig. 4.5: The three categories- Exposure, Sensitivity, and Resilience correspond to stressor-specific, overlap-specific, and habitat-specific criteria respectively.

5. **Optional** Determine whether spatial criteria are available and desired. These are vector layer files which would

provide more explicit detail for a specific criteria in the assessment. If spatially explicit criteria is desired, this input should point to an outer directory for all spatial criteria. A rigid structure **MUST** be followed in order for the model to run. Within the outer spatial criteria folder, there **MUST** be the following 3 folders: Sensitivity, Exposure, and Resilience. Vector criteria may then be placed within the desired folder. Each feature in the shapefiles used **MUST** include a 'Rating' attribute which maps to a float or int value desired for use as the rating value of that spatial criteria area.

- Any criteria placed within the Resilience folder will apply only to a given habitat. They should be named with the form: habitatname_criterianame.shp. Criteria names may contain more than one word if separated by an underscore.
- Any criteria placed within the Exposure folder will apply to the overlap between a given habitat and a given stressor. They should be named with the form: habitatname_stressorname_criterianame.shp. Criteria names may contain more than one word if separated by an underscore.
- Any criteria placed within the Sensitivity folder will apply to the overlap between a given habitat and a given stressor. They should be named with the form: habitatname_stressorname_criterianame.shp. Criteria names may contain more than one word if separated by an underscore.



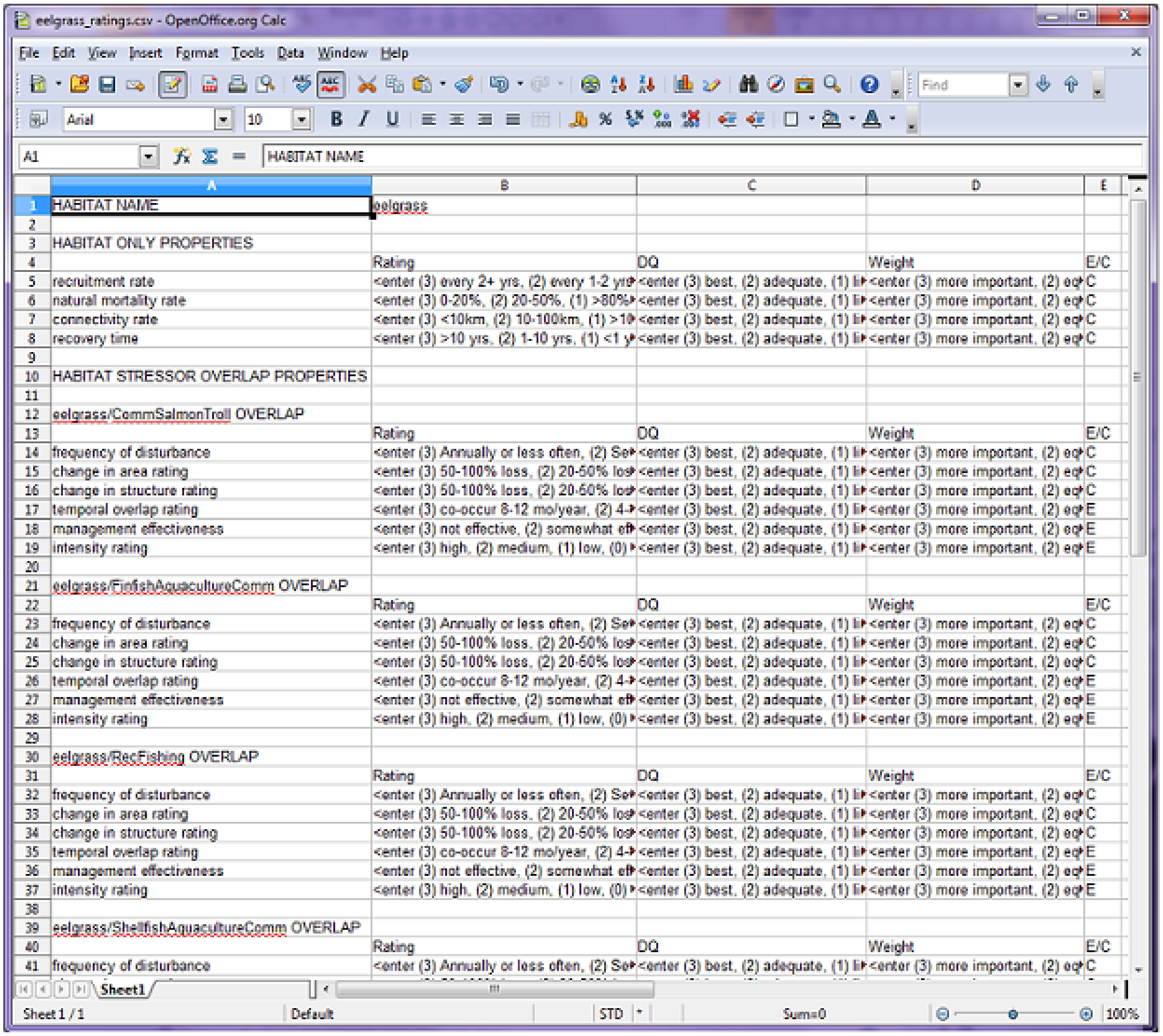
6. Run the tool. This will create a directory in your selected workspace called habitat_stressor_ratings. Keep in mind that if a folder of the name habitat_stressor_ratings already exists within the workspace, it will be deleted to make way for the new output folder. This directory can be renamed as necessary after completion, and will contain a series of files with the form: habitatname_ratings.csv, as well a file named stressor_buffers.csv. There will be one file for every habitat, and the one additional file for stressor buffers. HRA 3.0's sample data includes a sample folder for use within the main HRA executable called habitat_stressor_ratings_sample, containing pre-filled criteria values relevant to the sample data for the west coast of Vancouver Island, Canada.

Ratings CSVs

The CSVs contained within the habitat_stressor_ratings folder will provide all criteria information for the run of the Habitat Risk Assessment. There are two types of CSVs- habitat overlap CSVs and the stressor buffer CSV. Habitat CSVs will contain not only habitat-specific criteria information, but also all criteria that impact the overlap between that habitat and all applicable stressors. The stressor buffer CSV will be a single file containing the desired buffer for all stressors included in the assessment.

When preprocessor is run, the CSVs will contain no numerical ratings, only guidance on how each rating might be filled out. The user should use the best available data sources in order to obtain rating information. The column information to be filled out includes the following:

1. "Rating"- This is a measure of a criterion's impact on a particular habitat or stressor, with regards to the overall ecosystem. Data may come from a combination of peer-reviewed sources at the global scale and locally available fine-scale data sources. Model inputs and results can be updated as better information becomes available. We provide guidance for well-known criteria on a scale of 0-3, but it should be noted that if information is available on a different scale, this can also be used. It is important to note, however, that all rating information across all CSVs should be on one consistent scale, regardless of what the upper bound is.



	A	B	C	D	E
1	HABITAT NAME	eelgrass			
2					
3	HABITAT ONLY PROPERTIES				
4		Rating	DQ	Weight	E/C
5	recruitment rate	<enter (3) every 2+ yrs, (2) every 1-2 yrs	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	C
6	natural mortality rate	<enter (3) 0-20%, (2) 20-50%, (1) >80%	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	C
7	connectivity rate	<enter (3) <10km, (2) 10-100km, (1) >10	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	C
8	recovery time	<enter (3) >10 yrs, (2) 1-10 yrs, (1) <1 y	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	C
9					
10	HABITAT STRESSOR OVERLAP PROPERTIES				
11					
12	eelgrass/CommSalmonTroll OVERLAP				
13		Rating	DQ	Weight	E/C
14	frequency of disturbance	<enter (3) Annually or less often, (2) Se	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	C
15	change in area rating	<enter (3) 50-100% loss, (2) 20-50% los	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	C
16	change in structure rating	<enter (3) 50-100% loss, (2) 20-50% los	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	C
17	temporal overlap rating	<enter (3) co-occur 8-12 mo/year, (2) 4-	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	E
18	management effectiveness	<enter (3) not effective, (2) somewhat ef	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	E
19	intensity rating	<enter (3) high, (2) medium, (1) low, (0)	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	E
20					
21	eelgrass/FinfishAquacultureComm OVERLAP				
22		Rating	DQ	Weight	E/C
23	frequency of disturbance	<enter (3) Annually or less often, (2) Se	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	C
24	change in area rating	<enter (3) 50-100% loss, (2) 20-50% los	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	C
25	change in structure rating	<enter (3) 50-100% loss, (2) 20-50% los	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	C
26	temporal overlap rating	<enter (3) co-occur 8-12 mo/year, (2) 4-	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	E
27	management effectiveness	<enter (3) not effective, (2) somewhat ef	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	E
28	intensity rating	<enter (3) high, (2) medium, (1) low, (0)	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	E
29					
30	eelgrass/RecFishing OVERLAP				
31		Rating	DQ	Weight	E/C
32	frequency of disturbance	<enter (3) Annually or less often, (2) Se	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	C
33	change in area rating	<enter (3) 50-100% loss, (2) 20-50% los	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	C
34	change in structure rating	<enter (3) 50-100% loss, (2) 20-50% los	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	C
35	temporal overlap rating	<enter (3) co-occur 8-12 mo/year, (2) 4-	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	E
36	management effectiveness	<enter (3) not effective, (2) somewhat ef	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	E
37	intensity rating	<enter (3) high, (2) medium, (1) low, (0)	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	E
38					
39	eelgrass/ShellfishAquacultureComm OVERLAP				
40		Rating	DQ	Weight	E/C
41	frequency of disturbance	<enter (3) Annually or less often, (2) Se	<enter (3) best, (2) adequate, (1) li	<enter (3) more important, (2) eq	C

Fig. 4.6: Upon initial creation, CSVs will contain no ratings, only guidance for known criteria on a scale of 0-3. However, users should feel free to fill in ratings on a different scale if there is significant reviewed data, but should be sure to be consistent on scale across ALL CSVs.

2. "DQ"- This column represents the data quality of the score provided in the 'Rating' column. Here the model gives the user a chance to downweight less-reliable data sources, or upweight particularly well-studied criteria. While we provide guidance for a scoring system of 1-3, the user should feel free to use any upper bound they feel practical, as long as the scale is consistent. The lower bound, however, should ALWAYS be 1, unless the user wishes to remove the entire criteria score.
3. "Weight"- Here the user is given the opportunity to upweight critiera which they feel are particularly important to the system, independent of the source data quality. While we provide guidance for a scoring system from 1-3, the user should feel free to use any upper bound they feel practical, as long as the scale is consistent. The lower bound, however, should ALWAYS be 1 unless the user wishes to remove the entire criteria score.
4. (Optional) "E/C"- This column indicates whether the given criteria are being applied to the exposure or the consequence portion of the chosen risk equation. These can be manually changed by the user on a single criterion basis, however, we would strongly recommend against it. If the user desires to change that criterion's allocation, it would be better to change the allocation of the criterion within the Resilience, Exposure, Sensitivity categories using the HRA Preprocessor User Interface. By default, any criteria in the Sensitivity or Resilience categories will be assigned to Consequence (C) within the risk equations, and any criteria within the Exposure category will be assigned to Exposure (E) within the risk equation.

Note: Required ratings data - We recommend users include information about all of the key components of risk (i.e., spatial overlap and other exposure criteria, consequence criteria and the components of consequence, resilience and sensitivity. Nevertheless, the model will produce estimates for risk with only the habitat and stressor spatial layers and no other exposure values (i.e., E = 0 = no score for all other exposure criteria). To produce these estimates, the model does require values for at least one consequence criteria, either sensitivity or resilience. Without this information, the model will return an error message. If the user inputs scores for only sensitivity or resilience, then the consequence score will be based on those data alone.

Note: Specifying No Interaction Between Habitat and Stressor - As of InVEST 3.3.0 the HRA model will allow users to indicate that a habitat - stressor pair should have no interaction. This essentially means that the model will consider the habitat and stessor have no spatial overlap. This enhancement is to deal with the issue of having fine resolution vector data where the values may share the same pixel space when converted to a raster grid format. To set a habitat - stressor pair to no overlap, simply fill in each criterias "Rating" column with an "NA" value for the given pair. ALL "Rating" values for that pair must be set to "NA" for the model to consider the pair to have no interaction / overlap. If an "NA" value is found, but not all are set, an error message will be presented.

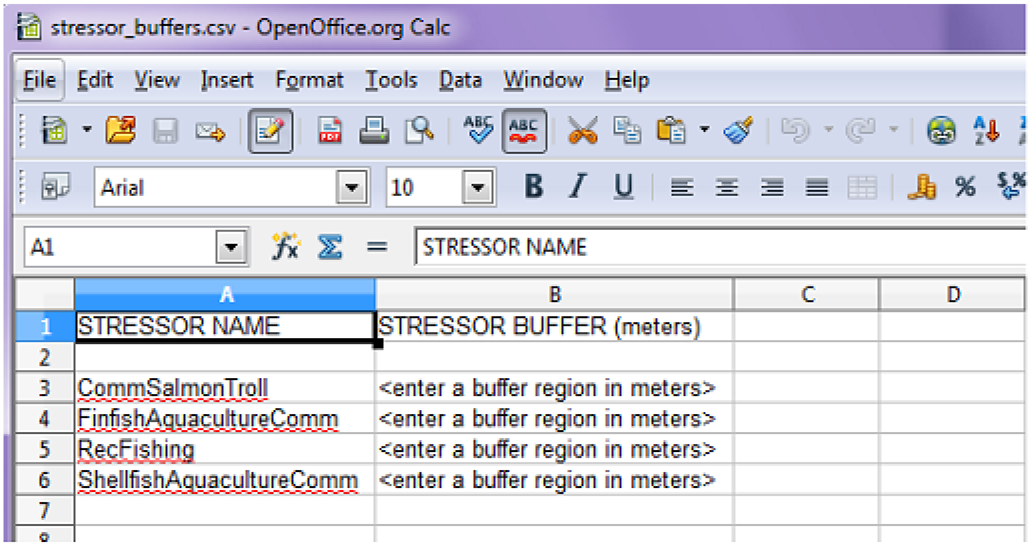
Habitat CSVs should be filled out with habitat-specific criteria information as well as any criteria which apply to the overlap of the given habitat and stressors. The Stressor Buffer CSV should be filled out with the desired numerical buffer which can be used to expand a given stressor's influence within the model run. This can be 0 if no buffering is desired for a given stressor, but may NOT be left blank.

Any criteria which use spatially explicit criteria (specified by the user during the HRA Preprocessor) will be noted in the CSV by the word 'SHAPE' in the rating column for that habitat, stressor, or combined criteria. The user should still fill in a Data Quality and Weight for these criteria, but should **NOT** remove the 'SHAPE' string unless they no longer desire to use a spatial criteria for that attribute.

Habitat Risk Assessment

The main computation portion of the HRA model will be done by the Habitat Risk Assessment executable. First we describe required inputs. The required inputs are the minimum data needed to run this model.

1. **Workspace Location (required)**. Users are required to specify a workspace folder path. It is recommended that the user create a new folder for each run of the model. For example, by creating a folder called "runBC" within the "HabitatRiskAssess_3_0" folder, the model will create "Intermediate" and "Output" folders within this "runBC" workspace. The "Intermediate" folder will compartmentalize data from intermediate processes.



The model's final outputs will be stored in the "Output" folder.

```
Name: Path to a workspace folder.  Avoid spaces.
Sample path: \InVEST\HabitatRiskAssess_3_0\runBC
```

2. **Criteria Scores Folder (required)**. After running the HRA Preprocessor tool, a folder will be created which contains the collective criteria scores for all habitats and stressors. For this input, point to the outer folder containing all CSVs.

```
Name: Folder can be named anything, but avoid spaces.
Sample path: \InVEST\HabitatRiskAssess_3_0\runBC\habitat_stressor_ratings
```

3. **Resolution of Analysis (required)**. The size in meters that is desired for the analysis of the shapefile layers at a grid cell scale. This will define the width and height of each unique risk grid cell. This must be a whole number. The user should base this size on the resolution of the habitat data and scale at which habitats are distributed in space. For example, small patches of seagrasses and kelp are often about 100-200 square meters, which is about the smallest resolution we recommend running the model. If the input habitat data are coarse, then a minimum of 500 meters is better. If you examine your risk outputs and find that the edges of patches of habitat have regular and distinct variation in risk, such that every high and medium risk cell on the edge of habitat patches are border by low risk cells, consider enlargening your resolution. We recommend running the model for the first time at a low resolution (500 m or 1 km) to verify that the model is running properly. Then use a higher resolution in subsequent runs.

4. **Risk Equation (required)**. This selection chooses the equation that will be used when calculating risk to a given habitat. (See the *Risk of Human Activities to Habitats* section.) The user may choose either either a Euclidean risk model, or a Multiplicative risk model.

5. **Decay Equation (required)** This selection influences how the "zone of influence" (i.e., buffer distance) of a stressor will be applied to risk. The stressor buffer distance in the stressor buffer CSV can be degraded to provide a more accurate depiction of the influence of a stressor beyond its footprint. The decay equation decays the overall exposure rating (e.g., combined spatial overlap, temporal overlap, intensity, management effectiveness) before the value for E goes into the risk equation. For each pixel, the model uses the value of the decayed exposure score. The options for decay are as follows. "None" will apply the full exposure to the full range of the stressor footprint plus buffer, without any decay. "Linear" and "Exponential" will use the stated equation as a model for decay from the edges of the footprint to the extent of the buffer distance.

6. **Maximum Criteria Score (required)** The maximum criteria score is the user-reported highest value assigned to any criteria rating within the assessment. This will be used as the upper bounded value against which all

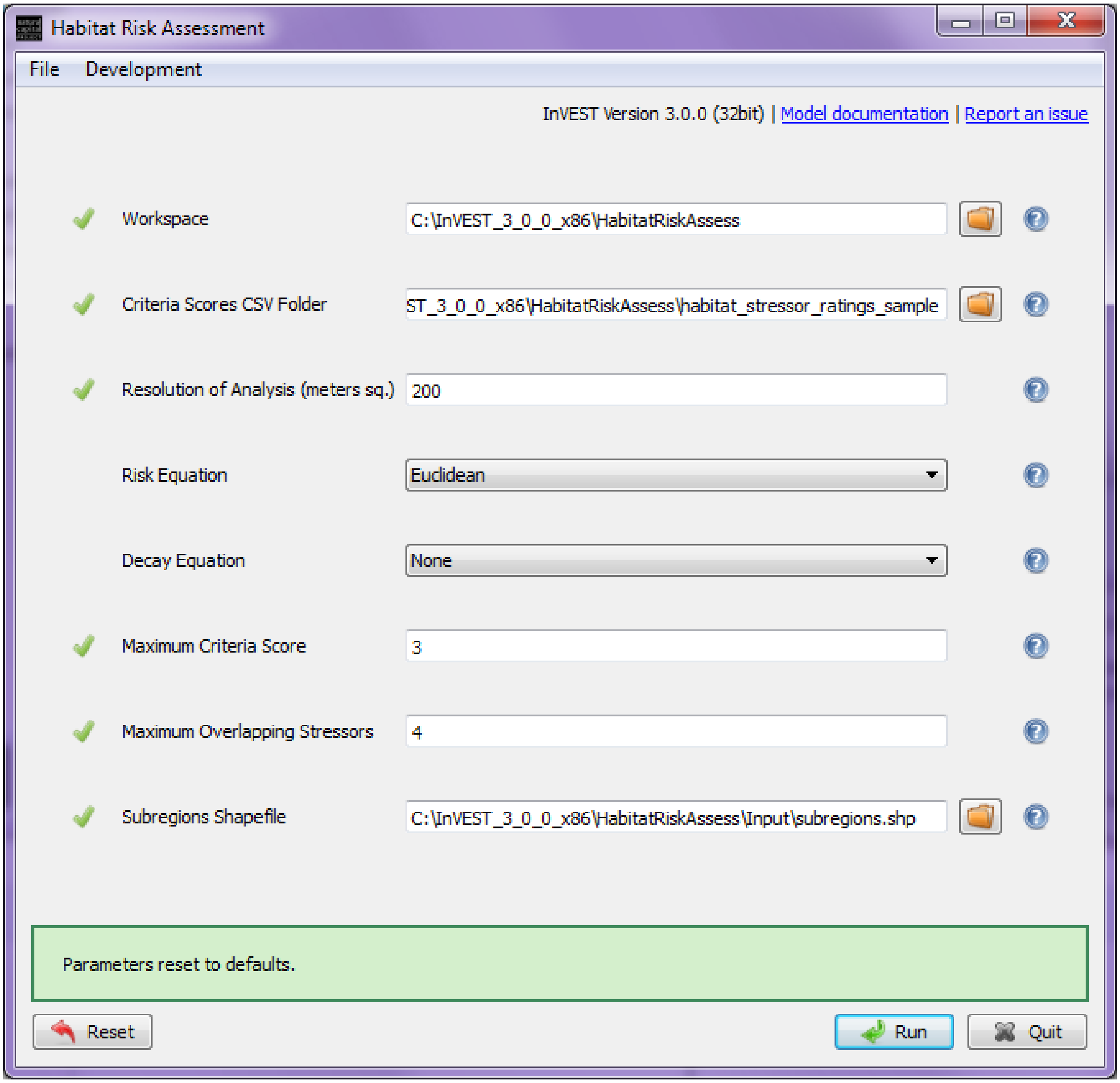


Fig. 4.7: The HRA 3.0 main executable.

rating scores will be compared. For example, in a model run where the ratings scores vary from 0-3, this would be a 3. If the user chooses to use a different scale for ratings, however, this should be the highest value that could be potentially assigned to a criteria. If the model run is using spatially explicit criteria, this value should be the maximum value assigned to either a criteria feature or to a CSV criteria rating.

7. **Maximum Overlapping Stressors (required)** The is the largest number of stressors that overlap withing the analysis zone. This will be used in order to make determinations of low, medium, and high risk for a given habitat. If the number of overlapping stressors provided is too low, results will likely show more medium and high risk areas than are present. Conversely, if the number of overlapping stressors is too high, it will be difficult for areas to break the threshold to show up as medium or high risk. If unsure how many stressors overlap, we recommend running the overlap analysis tool without weighting.

8. **Use Subregions Shapefile? (required)**. The model will use a subregions shapefile to generate an HTML table of averaged exposure, consequence, and risk values within each subregion by habitat and stressor. In addition, if the Risk Equation chosen is Euclidean, the model will also generate a series of figures which clearly display the exposure-consequence ratings and the resulting risk results for each habitat-stressor combination by subregion. It will also create a figure showing cumulative ecosystem risk for all subregions habitats in the study. Each of the subregion shapefile features **MUST contain a 'Name' attribute** in order to be properly included in the subregion averaging. If subregion data is not available for the given study region, an AOI for the area could also be used in order to obtain averaged data per habitat-stressor pair. However, the AOI must also contain a 'Name' attribute.:

```
Name: File can be named anything, but avoid spaces.
File Type: Polygon shapefile (.shp)
Sample path: \InVEST\HabitatRiskAssess_3_0\runBC\subregions.shp
```

4.2.5 Interpreting Results

Model Outputs

Upon successful completion of the model, you will see new folders in your Workspace called "Intermediate" and "Output". These two folders will hold all outputs, both temporary and final that are used in a complete run of the model. While most users will be interested only in the Output folder data, we will describe all outputs below.

Intermediate Folder

The Intermediate folder contains files that were used for final output calculations. All rasters within this file use the pixel size that the user specifies in the "Resolution of Analysis" text field of the *Habitat Risk Assessment* main executable.

- \Intermediate\Criteria_Rasters\spatial_criteria_name.tif
  - If the user has included any spatially explicit criteria in the assessment, this folder will contain a rasterized version of that vector layer, with the 'Rating' attribute burned as the pixel value.
- \Intermediate\Habitat_Rasters\habitat_name.tif
  - A rasterized version of all habitat or species vector files included in the assessment.
- \Intermediate\Stressor_Rasters\stressor_name.tif
  - A rasterized version of all stressor vector files included in the assessment.
- \Intermediate\Stressor_Rasters\stressor_name_buff.tif
  - This is a copy of the stressor_name.tif file in the same folder, but with each stressor's individual buffering included. If a given stressor has a 0 buffer distance, this will be an exact copy of the rasterized vector file.

For all other files, this will be buffered by the desired amount set forth in the "Stressor Buffer (m)" section of the *Ratings CSVs*, decayed from 1 to 0 using the equation chosen in the "Decay Equation" section of the *Habitat Risk Assessment*.

- \Intermediate\Overlap_Rasters\H[habitat_name]_S[stressor_name].tif
  - A raster representing the overlap between each pair of the habitat or species rasters, and the buffered stressor rasters.
- \Intermediate\H[habitatname]_S[stressorname]_C_Risk_Raster.tif
  - A raster representing the Consequence portion of the final risk calculations for the overlap of the given habitat and stressor.
- \Intermediate\H[habitatname]_S[stressorname]_E_Risk_Raster.tif
  - A raster representing the Exposure portion of the final risk calculations for the overlap of the given habitat and stressor.
- Intermediate\H[habitatname]_S[stressorname]_Risk_Raster.tif
  - A raster containing the final risk calculation for the given habitat and stressor combination. This risk raster takes into account each of the criteria that apply to the habitat and stressor, as well as the user-specified risk equation.

Output Folder

The following is a short description of each of the final outputs from the HRA model. Each of these output files is saved in the "Output" folder that is saved within the user-specified workspace directory:

GIS

- \Output\maps\recov_potent_H[habitat_name].tif
  - This raster layer depicts the recovery potential of each cell for the given habitat. Recovery potential is typically based on natural mortality rate, recruitment rate, age at maturity/recovery time and connectivity, though these can be altered by the user if alternate criteria are desired. Recovery potential is useful to those who are interested in identifying areas where habitats are more resilient to human stressors, and therefore may be able to withstand increasing stress. Habitats with low recovery potential are particularly vulnerable to intensifying human activities.
- \Output\maps\ecosys_risk.tif
  - This raster layer depicts the sum of all cumulative risk scores for all habitats in each grid cell. It is best interpreted as an integrative index of risk across all habitats in a grid cell. For example, in a nearshore grid cell that contains some coral reef, mangrove and soft bottom habitat, the ecosys_risk value reflects the risk to all three habitats in the cell. The "ecosys_risk" value increases as the number of habitats in a cell exposed to stressors increases.
- \Output\maps\cum_risk_H[habitat_name].tif
  - This raster layer depicts the cumulative risk for all the stressors in a grid cell on a habitat-by-habitat basis. For example, "cum_risk_eelgrass" depicts the risk from all stressors on habitat "eelgrass". Cumulative risk is derived by summing the risk scores from each stressor (i.e. more stressors leads to higher cumulative risk). This layer is informative for users who want to know how cumulative risk for a given habitat varies across a study region (e.g. identify hotspots where eelgrass or kelp is at high risk from multiple stressors). Hotspots of high cumulative risk may be targeted for restoration or monitoring.
- \Output\maps\[habitat_name]_RISK.shp

- These shapefiles are habitat specific and are classified by amount of risk. Each feature in the shapefile has a 'CLASSIFY' attribute, which will be 'LOW'/'MEDIUM'/'HIGH', depending on the amount of risk each contains relative to the risk thresholds. The thresholds of low/med/high are determined in one of two ways. A particular habitat pixel is considered high risk if any of the habitat-stressor risk pixels which make it up are > 66% of the total potential risk of any habitat-stressor pixel, or if the habitat risk map itself is > 66% of the total user-defined max potential risk (as determined by the maximum overlapping stressors within a habitat). Medium risk pixels use the same guidelines, but are defined by risk that falls between 33% and 66%. Low risk is any pixels below 33%. There is one habitat risk shapefile for each vector file originally used within the assessment.

HTML and Plots These outputs are optional, and their creation is dependent on user-provided subregion shapefiles.

- \Output\HTML_Plots\risk_plot_AOI[aoiname].html
  - These figures show the cumulative risk for each habitat within a given subregion. There will be one subplot for every habitat or species. Within the habitat plot, there are points for every stressor. Each point is graphed by Exposure, Consequence values. If the risk equation chosen was Euclidean, the distance from the stressor point to the origin represents the average risk for that habitat, stressor pair within the selected AOI. Stressors that have high exposure scores and high consequence scores pose the greatest risk to habitats. Reducing risk through management is likely to be more effective in situations where high risk is driven by high exposure, not high consequence.
- \Output\HTML_Plots\ecosystem_risk_plot.png
  - This figure shows the cumulative risk for each habitat in the study region by subregion. This figure can be used to determine which habitats are at highest risk from human activities, and if this risk is mostly due to high cumulative exposure (exogenous factors that can be mitigated by management) or high cumulative consequence (endogenous factors that are less responsive to human intervention).
- \Output\HTML_Plots\Sub_Region_Averaged_Results_[yr-mon-day-min-sec].html
  - This HTML table is a concatenated set of data for all pairings of habitat and stressor within each provided subregion. For every pairing of habitat and stressor, the table provides average exposure, consequence, risk, and risk percentage (as a portion of total potential risk).

Log File

- hra-log-yr-mon-day-min-sec.txt
  - Each time the model is run a text file will appear in the workspace folder. The file will list the parameter values for that run and be named according to the date and time.
  - Parameter log information can be used to identify detailed configurations of each of scenario simulation.

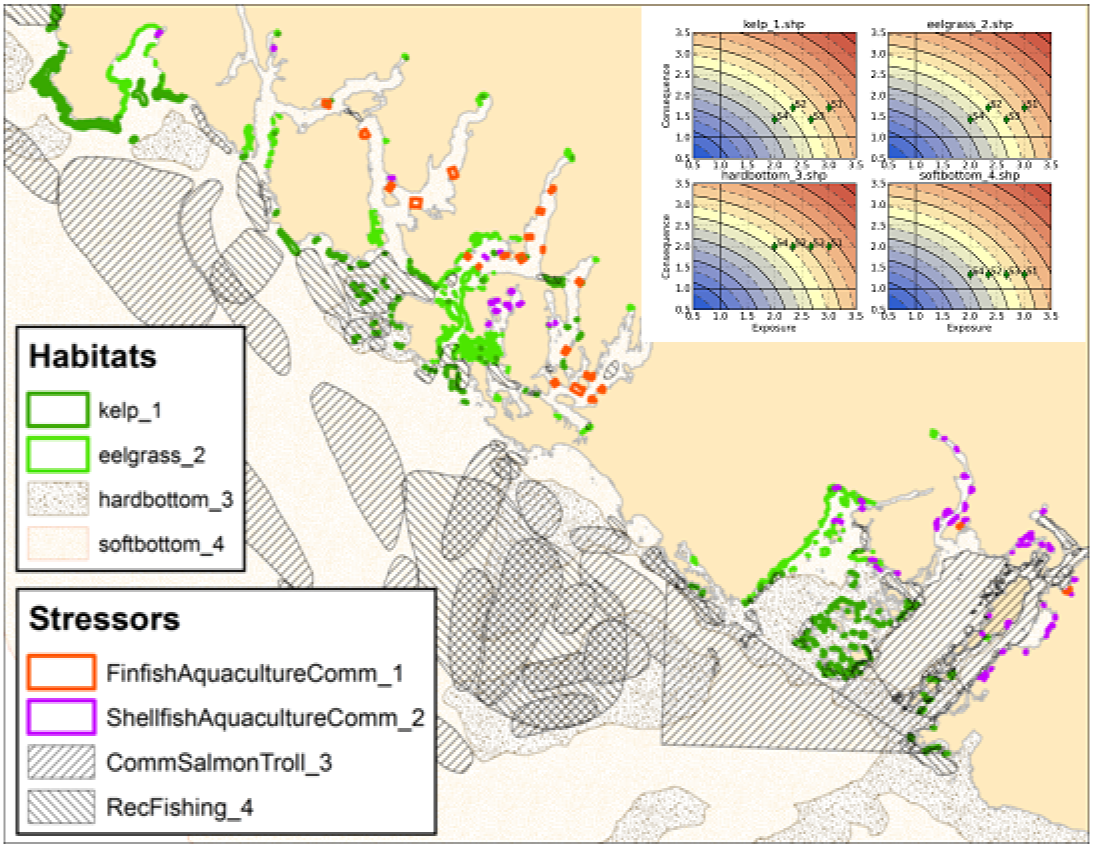
4.2.6 References

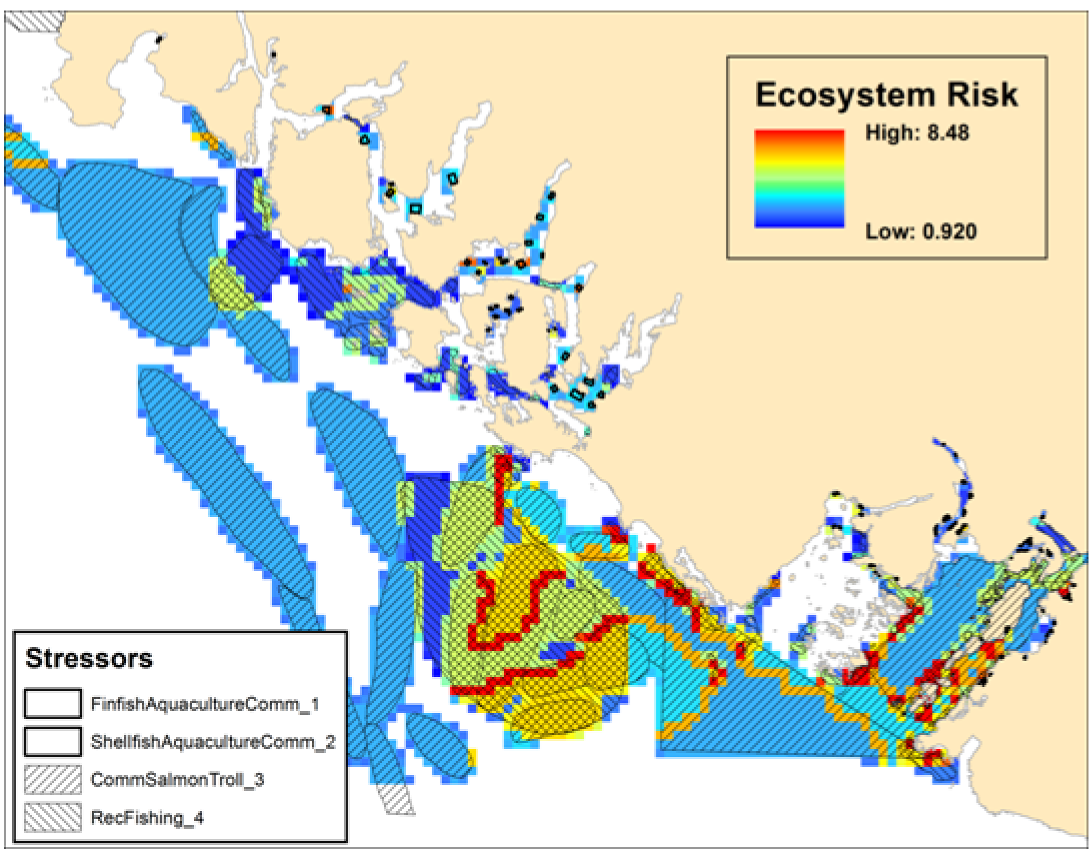
Astles, K. L., Holloway, M. G., Steffe, A., Green, M., Ganassin, C., & Gibbs, P. J. 2006. An ecological method for qualitative risk assessment and its use in the management of fisheries in New South Wales, Australia. Fisheries Research, 82: 290-303.

Burgman, M. 2005. Risks and decisions for conservation and environmental management. Cambridge University Press, Cambridge, UK.

Crain, C. M., Kroeker, K., & Halpern, B. S. 2008. Interactive and cumulative effects of multiple human stressors in marine systems. Ecology Letters, 11: 1304-1315.

Dawson, T. P., Jackson, S. T., House, J. I., Prentice, I. C., & Mace, G. M. 2011. Beyond Predictions: Biodiversity Conservation in a Changing Climate. Science, 332: 53-58.


kelp_1.shp
eelgrass_2.shp
hardbottom_3.shp
softbottom_4.shp
Consequence
Exposure
Habitats
kelp_1
eelgrass_2
hardbottom_3
softbottom_4
Stressors
FinfishAquacultureComm_1
ShellfishAquacultureComm_2
CommSalmonTroll_3
RecFishing_4


Ecosystem Risk
High: 8.48
Low: 0.920
Stressors
FinfishAquacultureComm_1
ShellfishAquacultureComm_2
CommSalmonTroll_3
RecFishing_4

Halpern, B. S., Walbridge, S., Selkoe, K. A., Kappel, C. V., Micheli, F., D'Agrosa, C., Bruno, J. F., et al. 2008. A Global Map of Human Impact on Marine Ecosystems. Science, 319: 948-952.

Hobday, A. J., Smith, A. D. M., Stobutzki, I. C., Bulman, C., Daley, R., Dambacher, J. M., Deng, R. A., et al. 2011. Ecological risk assessment for the effects of fishing. Fisheries Research, 108: 372-384.

Samhouri, J. F., and P. S. Levin. Linking Land- and Sea-Based Activities to Risk in Coastal Ecosystems. 2012. Biological Conservation 145(1): 118–129. doi:10.1016/j.biocon.2011.10.021.

Teck, S. J., Halpern, B. S., Kappel, C. V., Micheli, F., Selkoe, K. A., Crain, C. M., Martone, R., et al. 2010. Using expert judgement to estimate marine ecosystem vulnerability in the California Current. Ecological Applications 20: 1402-1416.

Williams, A., Dowdney, J., Smith, A. D. M., Hobday, A. J., & Fuller, M. 2011. Evaluating impacts of fishing on benthic habitats: A risk assessment framework applied to Australian fisheries. Fisheries Research, In Press.

4.3 Marine Water Quality

4.3.1 Summary

Management of water quality plays a critical role in human and ecosystem health in coastal and estuarine ecosystems. The lack of a predictive understanding of the dispersal and fate of contaminants is a major obstacle to the development of management strategies for water quality problems. We developed a marine water quality model consisting of physical transport and biogeochemical processes to simulate the dispersal of water quality state variables (e.g. contaminants) in response to changes in ecosystem structure driven by various management decisions and human activities. Hence, this model assesses how management and human activities influence the water quality in coastal and estuarine ecosystems. Although water quality is not an ecosystem service per se, the InVEST marine water quality model can be linked with other InVEST models to evaluate how changes in water quality might affect ecosystem services related to fisheries, aquaculture and recreation and how the exploitation of some services (e.g. aquaculture) might in turn affect water quality.

4.3.2 Introduction

The discharge of contaminants resulting from various management decisions and human activities may cause many types of water quality problems and potentially pose serious risks to both aquatic ecosystems and human health. Therefore, as human activities increase in coastal and marine ecosystems, water quality management has received increased attention in recent years. Since many processes (physical transport, biogeochemical and anthropogenic processes, etc.) affect water quality, it is difficult to determine the source of and to predict water quality problems. A numerical model based on physical and biogeochemical principals can help managers and decision makers investigate various water quality problems such as high concentrations of bacteria and toxic chemicals, hypoxia, and eutrophication (Park 1996).

Contaminants introduced into an estuarine system are transported by water movement (i.e. physical transport) and, while being transported, their concentrations are modified by biogeochemical processes. Therefore physical and biogeochemical processes combine to determine the fate of the contaminants. We developed a marine water quality model that accounts for both physical transport and biogeochemical processes to simulate the distribution and fate of a water quality state variable (e.g. contaminant or pollutant) in a coastal and estuarine system. The model allows users to change contaminant loadings from various sources, which may include sewage treatment plants, urban runoffs, storm sewers, failing septic systems, industrial discharges, floathomes, and aquaculture farms. For example, to explore the effects of alternate management schemes, users can alter pollutant or nutrient loading by adding, removing, or changing practices at aquaculture farms. They can also define pollutant or nutrient loading due to land based management.

The main output of the marine water quality model is a map of the concentration of a water quality state variable in response to the various management decisions under consideration. By exploring the concentration maps, users can assess—in a spatially explicit manner—how management and development strategies influence the water quality in their target area. The marine water quality model can be linked with other InVEST models to evaluate other ecosystem services related to fisheries, aquaculture, habitat quality, and recreation.

4.3.3 The Model

How it Works

The marine water quality model calculates the spatial distribution of water quality state variables by solving a tidal-average horizontal two-dimensional mass-balance equation.

$$E^T\left(\frac{\partial^2 C}{\partial x^2}+\frac{\partial^2 C}{\partial y^2}\right)-\left(U\frac{\partial C}{\partial x}+V\frac{\partial C}{\partial y}\right)+S=0 \tag{4.14}$$

Where

- x and y east and north coordinates, respectively
- C tidal averaged concentration of a water quality state variable
- U and V advective velocities (i.e., Eulerian residual current) in x and y directions, respectively
- E^T tidal dispersion coefficient
- S term to account for sources and sinks of pollutant

This is a steady state formulation of a classic advection diffusion equation. The first two terms on the left hand side represent tidal dispersion while the second two represent advective transport. The advective transport accounts for mass transport due to Eulerian residual current, which is obtained by averaging velocities at a fixed point over one or more tidal cycles. The tidal dispersion accounts for the dispersion of mass due to correlation between tidal components of velocity and concentration as well as the diffusion due to the turbulent fluctuations in velocity and concentration (MacCready & Geyer 2010). The tidal dispersion coefficient may be estimated by using observed salinity distribution or dye experiments. The observed tidal dispersion coefficient shows large variation ranging from 1 to 161 $\mathrm{km^2 day^{-1}}$ (Table 1) (Thomann & Mueller 1987).

Note that in a future version of this model E^T will be separated into a two dimensional vector with components Ex and E_y to correspond with each partial second order derivative in the first two terms of Equation (4.14).

Table 1. Tidal dispersion coefficient (E^T) in various estuarine systems (modified from Table 3.3 in Thomann & Mueller 1987).

Estuaries	Tidal dispersion coefficient ($\mathrm{km^2 day^{-1}}$)
Hudson River, NY	52
East River, NY	26
Wappinger and Fishkill Creek, NY	1-3
Delaware River, upper	5-18
Delaware River, lower	18-28
San Francisco Bay, southern	2-16
San Francisco Bay, northern	4-161
Rio Quayas, Ecuador	65
Thames River, England, low flow	5-7
Thames River, England, high flow	28

Tidal dispersion coefficient may also be parametrized as a function of tidal flow and length scale of an estuarine system (MacCready & Geyer 2010).

$$K = 0.035 \cdot U_T \cdot B \tag{4.15}$$

where

- U_T amplitude of depth-averaged tidal flow
- B the length scale of an estuarine system, which is the smaller of channel width or tidal excursion.

If users have tidal flow information, Equation (4.15) is a practical option to estimate spatially explicit tidal dispersion coefficient for their study area. Advective transport and tidal dispersion combine to determine physical transport of a water quality state variable. Physical transport processes take the same mathematical forms for all water quality state variables. That is, physical transport processes do not depend on the nature of the substances as long as the substances do not affect the water movement.

The last term (S) in Equation (4.14) represent material-specific biogeochemical processes consisting of internal sources/sinks (SI) and external sources/sinks (SE). SI is primarily due to the kinetic processes and SE includes pollutant loading into and removal from a water body. Different water quality state variables are affected by different biogeochemical processes and require appropriate kinetic formulations for each of the source and sink terms (Park 1996).

Biogeochemical Processes

Unlike physical transport processes, each water quality state variable is determined by different biogeochemical processes and requires appropriate kinetic formulations (Park 1996). The kinetic formulations are mostly empirical and thus have to be refined with the advances in our understanding of the representing kinetic processes. The InVEST marine water quality model provides users with a flexible framework to update or add biogeochemical processes for their target materials. An example of biogeochemical processes for pathogen simulation is given below.

Pathogens

Pathogens are disease-causing microorganisms that include bacteria, viruses, and protozoa, and can originate from many sources including sewage treatment plants, urban runoff, storm sewers, failing septic systems, industrial discharges, and contaminated sediments (Ji 2008). Contaminated water by pathogens is responsible for the spread of many contagious diseases, and understanding the dispersal and fate of pathogens is one of main concerns for water quality management. The studies of pathogens in surface water usually focus on indicator organisms such as fecal coliforms, E. coli or enterococci, and often consider a simple decay for the biogeochemical processes (Thomann & Mueller 1987):

$$S = -K_B C + \frac{W}{VOL} \tag{4.16}$$

where

- C concentration of indicator organism ($\mathrm{organism\ count\ m^{-3}}$)
- K_B decay rate ($\mathrm{day^{-1}}$)
- W external load of indicator organism ($\mathrm{organism\ count\ day^{-1}}$)
- VOL volume of water cell ($\mathrm{m^3}$)

As shown in Table 2, the average decay rate of total coliform bacteria is about 1.4 $\mathrm{day^{-1}}$ in freshwater (20°C) and 48 $\mathrm{day^{-1}}$ in seawater, but the maximum decay rate can be as large as 84 $\mathrm{day^{-1}}$ under optimal environmental conditions.

Table 2. Observed decay rates of indicator organisms (modified from Table 5.9 in Thomann & Mueller 1987).

Indictor organisms	$K_B(\mathrm{day}^{-1})$	Note
Total coliform	0.7-3.0 (avg. 1.4)	Average freshwater (20°C)
	8.0-84.0 (avg. 48.0)	Seawater (20°C) (variable temperature)
Total or fecal	0.0-2.4	New York Harbor Salinity: 2-18 $^0/_{00}$ (dark)
	2.5-6.1	New York Harbor Salinity: 15 $^0/_{00}$ (sunlight)
Fecal coliform	37.0-110.0	Seawater (sunlight)
E-Coli	0.08-2.0	Seawater, 10-30 $^0/_{00}$
Salmonella	0.1 - 3.0	Stormwater (20°C), Hamilton Bay (18°C)

Mancini (1978) made an equation to estimate decay rates of indicator bacteria as a function of salinity, temperature, sunlight and sink/resuspension.

$$K_B = [0.8 + 0.006(\%\ \text{sea water})]1.07^{(T-20)} + \frac{\alpha I_0}{K_e H}[1 - exp(-K_e H)] \pm \frac{v_s}{H} \tag{4.17}$$

Where

- T water temperature ($^\circ$C)
- α sunlight coefficient
- I_0 average solar radiation ($\mathrm{cal\ cm}^{-2}$)
- K_e light extinction coefficient (m^{-1})
- H average depth (m)
- v_s sink or resuspension rate ($\mathrm{m\ day}^{-1}$)

Users may consult Table 2 to find an appropriate K_B for their application. If users have enough data for the environmental conditions (water temperature, salinity, light information, etc.), Equation (4.17) may be applied to estimate K_B.

Boundary Condition

We need to define ocean and land boundary conditions to solve Equation (4.14) numerically. The ocean boundary (i.e., open boundary) indicates the outer boundary of the modeling domain adjacent to oceans. We assume the horizontal pollutant profile (e.g. C) is advected out of the modeling domain as a “frozen pattern”. That is

$$\nabla \cdot C_b = \nabla \cdot C_{b'} \tag{4.18}$$

where

- C_b is the concentration (organism count m^{-3}) on an inner boundary point b
- $C_{b'}$ is the concentration (organism count m^{-3}) on the outer boundary adjacent to point b

Additionally, no transport of C is allowed from or into the land.

Numerical Solution

We solve Equation (4.14) by using first and second order central difference expansions of the derivative terms and deriving an implicit Crank-Nicolson scheme. This scheme is unconditionally stable and has a truncation error of $O(\Delta h^2)$ where h is the discrete grid cell size.

4.3.4 Limitations and Simplifications

1. **Assumes a steady state condition (no time variation of model results)**: The current version assumes a steady state condition and cannot produce time varying model outputs. So, users should be aware that the model produces a distribution of a water quality state variable once it reaches an equilibrium status under defined conditions.
2. **Grid size of a water cell**: A finer grid size better resolves spatial differences in model outputs. However, it requires more computation and memory. Too many grid cells may cause an out of memory error.
3. **The credibility of physical transport**: Reliable information on physical transport processes is critical for reasonable model results.
4. **No vertical transport**: The governing mass balance equation of the model considers only horizontal transport of mass; it simulates vertically averaged conditions.
5. **Size of the modeling domain**: If the ocean boundary is too close to the pollutant loading points, inaccurate boundary values may artificially affect the model results. We therefore recommend users to prepare the grid system such that the ocean boundary is not too close to the loading locations.

4.3.5 Data Needs

The following are the data needs for the Marine Water Quality Model. The model is distributed with default arguments which are defaulted in the following parameters on the tool's first run.

- **Workspace**: The directory to hold output and intermediate results of the particular model run. After the model run is completed the output will be located in this directory. To run multiple scenarios, create a new workspace for each scenario.
- **Area of Interest (AOI)**: An ESRI Shapefile that contains a polygon indicating the target area. The output raster will align with the area of extents of this polygon. The polygon should be projected.
- **Land Polygon**: An ESRI Shapefile that contains a polygon indicating where the landmass lies. It should be in the same projection as the AOI polygon.
- **Output pixel size in meters**: Horizontal grid size, which determines the output resolution of the pollutant density raster. A larger number will make the output grid coarser but the model will run faster, while a finer resolution will require more computation and memory. Try making this number larger if a model run encounters an out of memory error.
- **Grid Cell Depth**: Grid size in a vertical direction (m), which is the layer thickness of the horizontal grid system.
- **Source Point Centroids**: An ESRI Shapefile that contains a point layer indicating the centroids of point pollutant sources that must have a field called Id that indicates the unique identification number for that point. This file must be in the same projection as the AOI polygon.
- **Source Point Loading Table**: Point source loading ($\mathrm{g\ day^{-1}}$ or $\mathrm{organism\ count\ day^{-1}}$) at the loading points that contains at least the headers ID and WPS which correspond to the identification number in the Source Point Centroids shapefile and the loading of pollutant at that point source.
- **Decay Coefficient (KB)**: Decay rate in the unit of $\mathrm{day^{-1}}$. Users may consult Table 2 or use Equation (4.17) to estimate K_B.
- **Dispersion Coefficients (E^T):** An ESRI Shapefile that contains a point layer with a field named `E_km2_day` indicating the dispersion coefficient ($\mathrm{km^2 day^{-1}}$) at that point as referenced in Equation (4.14). This file must be in the same projection as the AOI polygon. In a future release of this model this parameter will have x and y components.
- **(Optional) Advection Vectors (UV as point data):** An ESRI Shapefile that contains a point layer with two fields named `U_m_sec_` and `V_m_sec_` which correspond to the U and V components ($\mathrm{m\,s^{-1}}$) of the 2D

advective velocity vector as referenced in Equation (4.14). This file must be in the same projection as the AOI polygon.

4.3.6 Running the Model

To run the marine water quality model double click *invest_marine_water_quality_biophysical.exe* located in the folder entitled *invest-3* in the InVEST installation directory. The main interface indicates the required and optional input arguments as described in the **Data Needs** section above. Click the *Run* button to start the model. A successful run will be indicated in the window and a file explorer will open containing the results.

If you encounter any issues please post to the user's support forum at http://ncp-yamato.stanford.edu/natcapforums.

4.3.7 Interpreting Results

Model Outputs

Each of model output files is saved in the `Output` and `Intermediate` folders that are saved within the user-specified workspace.

Output folder

- `Output\concentration.tif`: The output raster indicating the concentration of the water quality state variable (e.g. pollutant). The units of the loading determine the units of output concentration. If the loading is expressed as the number (mass) of a targeted organism, the concentration unit is the number of organisms (mass) per volume.

Intermediate folder

- `Intermediate\in_water.tif`: This is a raster file indicating the land and water points that's used in the calculation of the domain to discritize Equation (4.14).
- `Intermediate\tide_e.tif`: This is a raster file with the interpolated values of the Tidal Dispersion Coefficients shapefile that are used to determine the ET values for each discritized grid cell.
- `Intermediate\adv_u.tif` and `intermediate\adv_v.tif`: This is a raster with the interpolated values of the Advection Vectors (UV as point data) shapefile that are used to determine the U and V components respectively.

4.3.8 Case Example Illustrating Model Inputs and Results

Managers and stakeholders want to estimate the distribution of fecal-coliform bacteria released from floathomes (recreational floating cabins, usually with untreated wastes) in sheltered areas along the west coast of Vancouver Island, BC, Canada. We have explored scenarios involving different levels of treatment (removal of fecal-coliform and thus a decreased loading) and different spatial arrangements of floathomes. Figs. 1 and 2 show a status quo arrangement of floathomes in Lemmens Inlet (and, in the case of Fig. 1, the surrounding area). We used an initial assumption that the loading of the untreated wastes from the floathomes of 1 million bacteria per day. In another scenario assumption, we modeled the effects of secondary treatment of waste from two floathomes (the 23rd and 24th in Fig. 2), assuming 95% removal (thus the initial loading is 50,000 bacteria per day). Model results, i.e. the distribution of fecal-coliform bacteria given the location of floathomes shown in Fig. 2 and the modeled treatment of waste described above, are shown in Fig. 3.

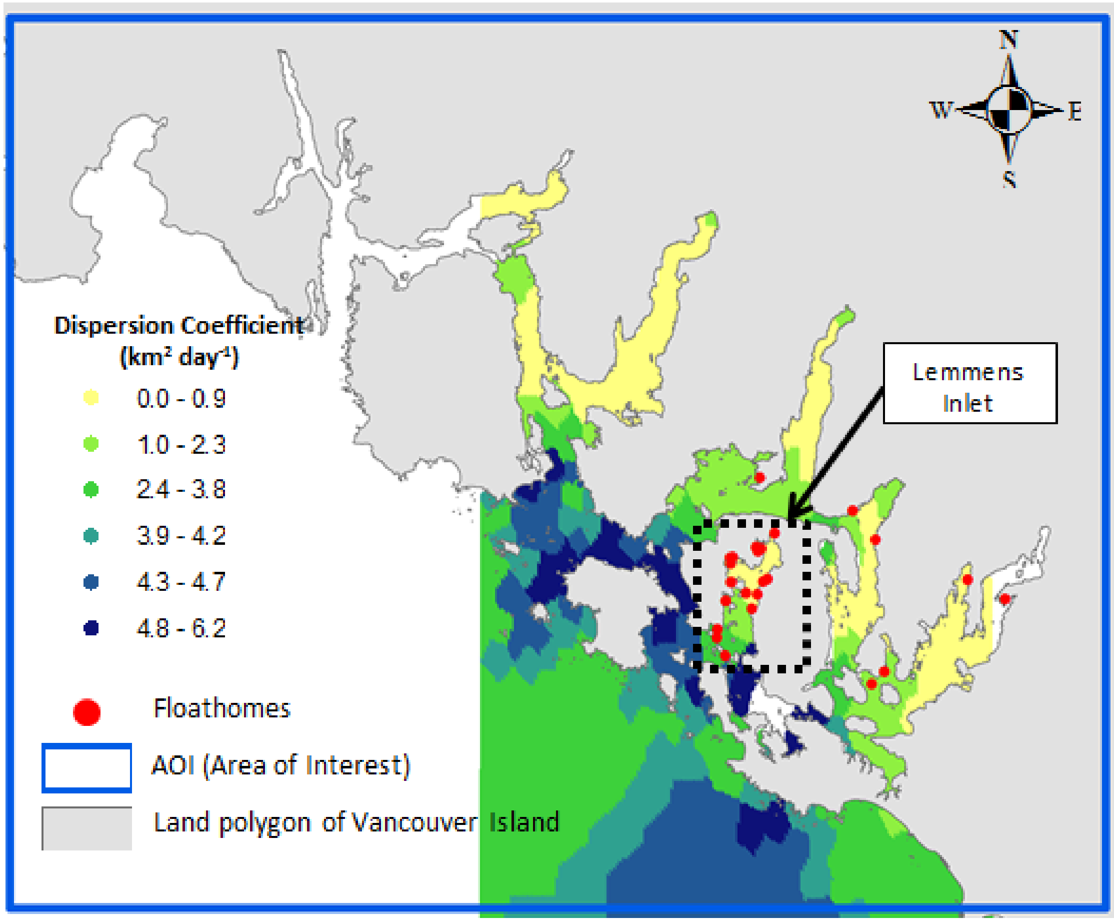


Figure 1. A map of Clayoquot Sound, BC, Canada showing a status quo arrangement of floathomes (red dots). The dotted box indicates Lemmens Inlet, the region of interest for potentially rearranging floathomes and/or exploring the effects of treating wastes. Background colors indicate tidal dispersion coefficients for the region, a key model input.

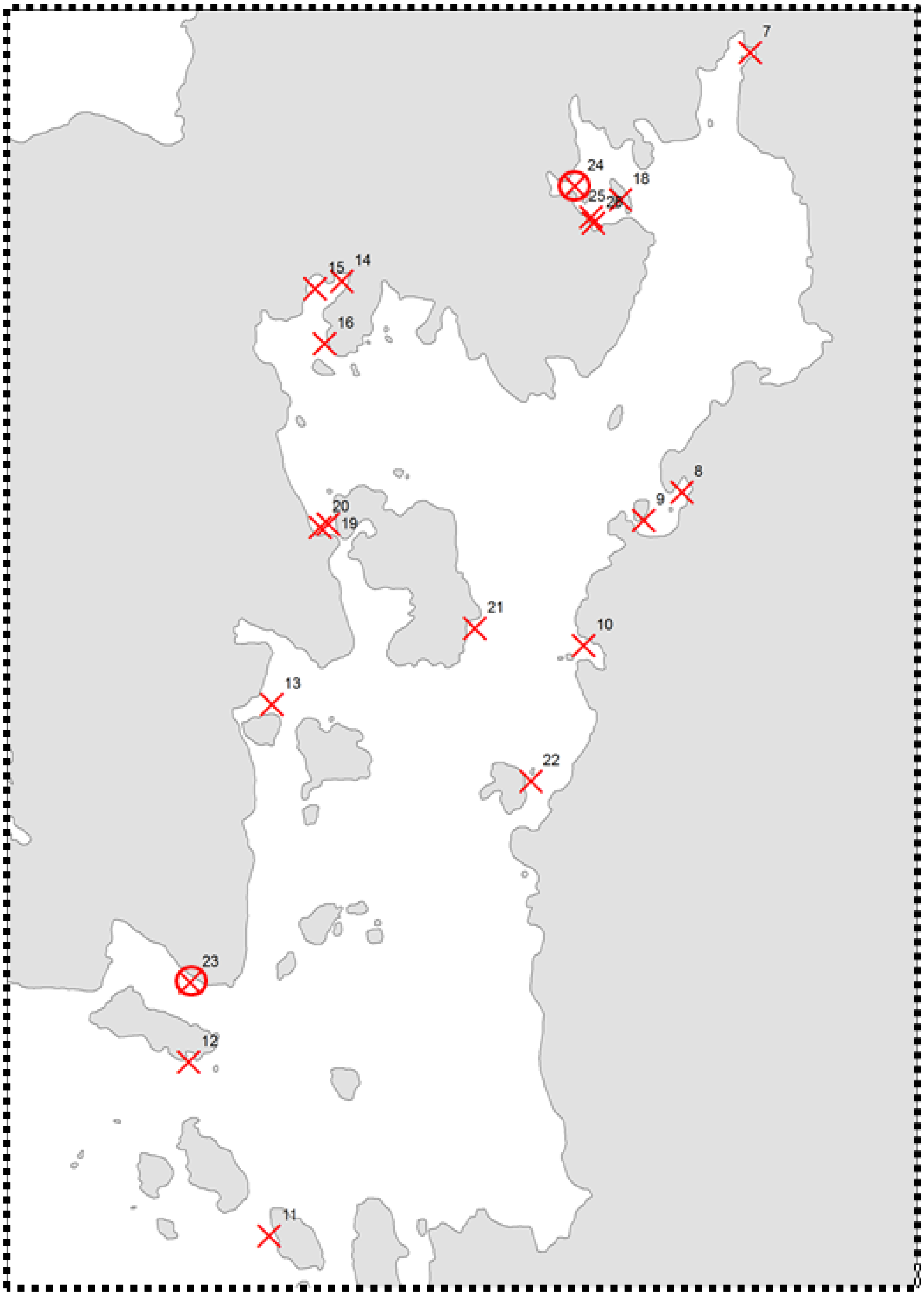


Figure 2. Enlarged map of Lemmens Inlet, showing the location of floathomes. Source point centroids are shown with red x's and red circles indicate treated wastes (23 and 24) assuming 95% removal of bacteria.

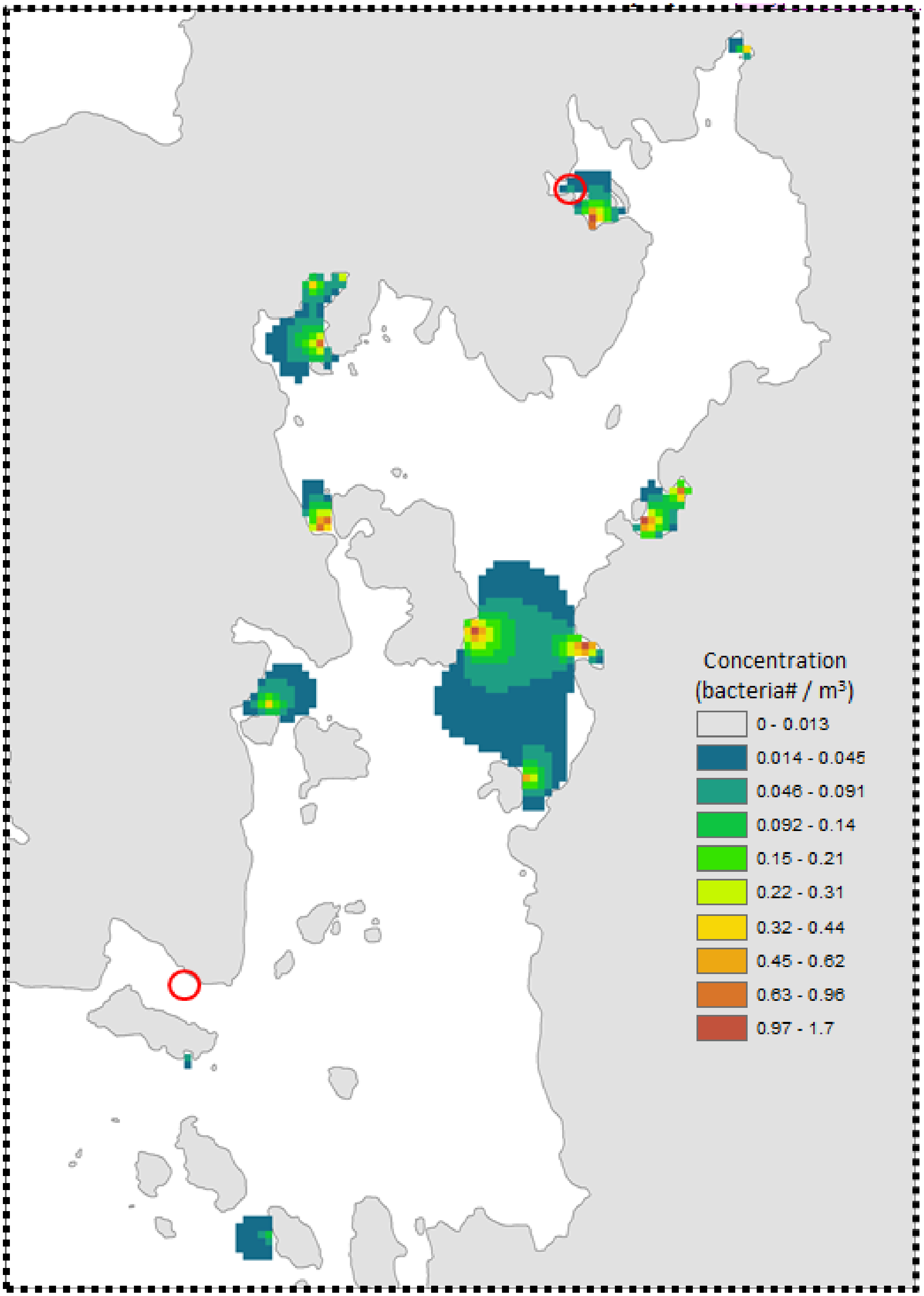


Figure 3. Map of modeled concentration of fecal coliform bacteria in Lemmens Inlet. Red circles indicate treated wastes. The results are for demonstration purposes only.

4.3.9 References

Park, K. 1996. Concept of surface water quality modeling in tidal rivers and estuaries. Environ. Eng. Res. 1:1-13.

MacCready, P., and W. R. Geyer 2010. Advances in estuarine physics. The Annual Review of Marine Science 2:35-58.

Thomann, R. V., and J. A. Mueller. 1987. Principles of surface water quality modeling and control. Prentice-Hall, NY.

Ji Z-G (2008) Hydrodynamics and water quality: Modeling rivers, lakes, and estuaries. New Jersey: John Wiley & Sons, Inc.

4.4 Pollinator Abundance: Crop Pollination

4.4.1 Summary

Seventy-five percent of globally important crops rely either in part or completely on animal pollination. The InVEST pollination model focuses on wild bees as a key animal pollinator. It uses estimates of the availability of nest sites and floral resources and bee flight ranges to derive an index of bee abundance nesting on each cell on a landscape (i.e., pollinator supply). It then uses flight range information to estimate an index of bee abundance visiting each agricultural cell. If desired, the model then calculates a simple index of the value of these bees to agricultural production, and attributes this value back to source cells. The results can be used to optimize agriculture and conservation investments. Required inputs include a current land use and land cover map, land cover attributes, species of pollinators present, and their flight ranges. The model's limitations include exclusion of non-farm habitats that may determine pollinator abundance and of the effects of land parcel size. The model also does not account for managed pollinators and pollinator persistence over time.

4.4.2 Introduction

Crop pollination by bees and other animals is a potentially valuable ecosystem service in many landscapes of mixed agricultural and natural habitats (Allen-Wardell et al. 1998, Free 1993). Pollination can increase the yield, quality, and stability of fruit and seed crops as diverse as tomato, canola, watermelon, coffee, sunflower, almond, and cacao. Indeed, Klein et al. (2007) found that 87 of 115 globally important crops benefit from animal pollination, a service valued variously in the billions to tens of billions per year globally (Costanza et al. 1997, Losey and Vaughan 2006, Nabhan and Buchmann 1997, Southwick and Southwick 1992).

Despite these numbers, it is important to realize that not all crops need animal pollination. Some crop plants are wind (e.g., staple grains such as rice, corn, wheat) or self pollinated (e.g., lentils and other beans), needing no animal pollinators to successfully produce fruits or seeds. Klein et al. (2007) provides a list of crops and their pollination requirements that can help identify whether crops in a region of interest may benefit from wild animal pollinators.

Decision-makers can use information on crop pollinators, their abundance across a landscape, and the pollination services they provide to crops in several ways. First, with maps of pollinator abundance and crops that need them, land use planners could predict consequences of different policies on pollination services and income to farmers (for an example, see Priess et al. 2007). Second, farmers could use these maps to locate crops intelligently, given their pollination requirements and predictions of pollinator availability. Third, conservation organizations or land trusts could use the tool to optimize conservation investments that benefit both biodiversity and farmers. Finally, governments or others proposing payment schemes for ecosystem services could use the results to estimate who should pay whom, and how much.

4.4.3 The Model

A wide range of animals can be important pollinators (e.g., birds, bats, moths and flies), but bees are the most important group for most crops (Free 1993). As a result, the InVEST Pollination model focuses on the resource needs and flight behaviors of wild bees. Many people think of honeybees, managed in artificial hives, when they think of pollinators, but wild bees also contribute to crop pollination. In fact, for several important crops (e.g., blueberries), native species are more efficient and effective pollinators than honeybees (Cane 1997). These native bees, in addition to feral honeybees living in the wild, can benefit crops without active management of captive hives. This is the pollination service associated with habitat conservation.

For bees to persist on a landscape, they need two things: suitable places to nest, and sufficient food (provided by flowers) near their nesting sites. If provided these resources, pollinators are available to fly to nearby crops and pollinate them as they collect nectar and pollen. The model therefore uses information on the availability of nesting sites and flower resources, as well as flight ranges of bees, to map an index of bee abundance across the landscape. In a second step, the model uses this map and bee flight ranges again to predict an index of the number of pollinators likely visiting crops in each agricultural cell on the landscape. If you opt to also estimate value indices, the model then takes a third and fourth step. In the third step, it uses a simplified yield function to translate bee abundance into crop value on each agricultural cell. And in the fourth step, it attributes these cell values back to cells "supplying" these bees. These steps are laid out in more detail below, and the full model description can be found in Lonsdorf et al. (in press).

How it Works

The model is based on a land use and land cover (LULC) map, showing both natural and managed land types. This map is divided into a regular grid of square cells, each of which is assigned a single LULC type. For each type, the model requires estimates of both nesting site availability and flower availability (e.g., for bee food: nectar and pollen). These data can be supplied from quantitative field estimates or from expert opinion, and are expressed in the form of relative indices (between 0 and 1). Flower availability can be supplied separately for different seasons if important, and the availability of nesting substrates can be estimated separately for multiple nesting guilds (e.g., ground nesters, cavity nesters).

Because bees are proficient flyers, they integrate over several elements of a landscape, moving between nesting habitats and foraging habitats (Ricketts et al. 2006). The distances they typically fly affect both their persistence and the level of service they deliver to farms. The model therefore requires a typical foraging distance for each pollinator species. These data can be supplied from quantitative field estimates (e.g., Roubik and Aluja 1983), proxies such as body size (Greenleaf et al. 2007), or from expert opinion.

Using these data, the model first estimates the abundance index of each pollinator species in every cell in the landscape, based on the available nesting sites in that cell and the flowers (i.e., food) in surrounding cells. Flowers in nearby cells are given more weight than distant cells, according to the species' average foraging range. Since pollinator abundance is limited by both nesting and floral resources, the pollinator abundance for species β index on cell x, $P_{x\beta}$, is the product of foraging and nesting such that:

$$P_{x\beta} = N_j \frac{\sum_{m=1}^{M} F_{jm} e^{\frac{-D_{mx}}{\alpha_\beta}}}{\sum_{m=1}^{M} e^{\frac{-D_{mx}}{\alpha_\beta}}}$$

where N_j is the suitability of nesting of LULC type *j*, F_j is the relative amount floral resources produced by LULC type j, Dmx is the Euclidean distance between cells m and x and α_β is the expected foraging distance for the pollinator β (Greenleaf et al. 2007).

The result is a map of the abundance index (0-1) for each species, which represents a map of "pollinator supply" (i.e., bees available to pollinate crops). In this sense, this map represents the potential sources of pollination services, but it has not yet incorporated demand. In other words, the landscape may be rich in pollinator abundance, but if there are no bee-pollinated crops on that landscape, those bees will not be providing the service of crop pollination.

To make this connection between areas of "supply" and "demand," the model calculates an abundance index of visiting bees at each agricultural cell, by again using flight ranges of pollinator species to simulate their foraging in nearby cells. Specifically, it sums pollinator supply values in cells surrounding each agricultural cell, again giving more weight to nearby cells. This sum, created separately for each pollinator species at each agricultural site, is an index of the abundance of bees visiting each farm site (i.e., "farm abundance"). We use the foraging framework described in the previous equation to determine the relative abundance of bees that travel from a single source cell x to forage on a crop in agricultural cell o:

$$P_{ox\beta} = \frac{P_{x\beta}e^{\frac{-D_{ox}}{\alpha_\beta}}}{\sum_{x=1}^{M} e^{\frac{-D_{ox}}{\alpha_\beta}}}$$

where $P_{x\beta}$ is the supply of pollinators of species β on cell x, $D_{ox\beta}$ is distance between source cell x and agricultural cell o for that species, and α_β is species' β average foraging distance. The numerator of this equation represents the distance-weighted proportion of the pollinators supplied by cell m that forage within cell o and the numerator is a scalar that normalizes this contribution by the total area within foraging distance (Winfree et al. 2005). The total pollinator abundance on agricultural cell o, Po, is simply the sum over all M cells. This second map represents the relative degree of pollination service at the demand points, or points at which this service is "delivered": agricultural cells.

The actual economic benefit received from pollination depends on how crops grown in each cell respond to pollinators. The model therefore takes two additional (optional) steps to translate farm abundances of pollinators into indices of expected economic value. In lieu of a more detailed agricultural production function, we use a simple saturating crop yield function, which assumes that yield increases as pollinator visitation increases, but with diminishing returns (Greenleaf and Kremen 2006). Crops vary in their dependence on pollinators; some crop species are self-compatible and yield is less dependent on pollination while other species obligately require pollination to generate any yield (Klein et al. 2007). We account for both observations, and thus calculate the expected yield of a crop c on farm o, $Y_{o\beta}$, as:

$$Y_{o\beta} = 1 - \nu_c + \nu_c \frac{P_{o\beta}}{P_{o\beta} + \kappa_c}$$

Where ν_c represents the proportion of total crop c's yield attributed only to wild pollination (e.g. ν_c would be equal to 1 if a crop is an obligately outcrossing species and equal to 0 if the crop species were wind-pollinated). In the denominator of the third term, κ_c is a half-saturation constant and represents the abundance of pollinators required to reach 50% of pollinator-dependent yield.

Once the model has calculated value for each agricultural cell, it redistributes this value back to cells that supplied the relevant pollinators, creating a map of value at the source. First, the model assigns fractions of the cell's value to each of the bee species, according to their partial contribution to total farm abundance. Then each species' value is redistributed back to the source cells from which they came using the same distance-weighted relationship described above. Thus source habitats close by provide greater service value than those farther away. Formally, we calculate pollinator service provided to O farms from each *m* cell, PS_m, as:

$$PS_{x\beta} = \nu_o P_{x\beta} \frac{\sum_{m=1}^{M} \frac{Y_{o\beta m}}{P_{o\beta m}} e^{\frac{-D_{mx}}{\alpha_\beta}}}{\sum_{m=1}^{M} e^{\frac{-D_{mx}}{\alpha_\beta}}}$$

where V_o represents the crop value in farm cell o. The result is a map of "pollinator service value" that estimates the relative index of economic value of pollinators for agricultural areas.

If the simple saturating yield function is deemed too simplistic, one may link this pollination model to InVEST's agricultural production model that includes other factors such as fertilizer, irrigation, labor, etc. The integration of these two models will give a more appropriate representation of the multiple inputs to agricultural production. It will also be possible to more specifically derive the amount of crop yield provided by wild pollinators (yield contribution) and the net present value of that additional yield. See Lonsdorf et al. (2009) and Lonsdorf et al. (in press) for equations that determine the pollinator supply, farm abundance, and pollinator service value maps.

Limitations and Simplifications

The model predicts an abundance index of wild pollinators on agricultural fields (cells) within a landscape, based on the pattern of land cover types and the resources they are estimated to contain for bees. It also converts this abundance into indices of production value and attributes this value to the source cells for pollinators. Like other InVEST models, the Pollination model is extremely simple, but it makes reasonably accurate predictions when compared to field observations of pollinators (Lonsdorf et al. 2009). Nevertheless, with this simplicity come several limitations that must be kept in mind.

First, the model predicts only relative patterns of pollinator abundance and pollination value (using indices of 0-1). This is because absolute estimates of nest density, resource availability, and pollinator abundance are rarely available, and yield functions (including pollinator abundance) for many crops are poorly defined. However, relying on relative indices limits our ability to estimate absolute economic values to better inform land-use planning decision-making, often based on cost-benefit analyses.

This simplicity is perhaps most limiting in calculating indices of value, both on farms and at the source cells of pollinator supply. With field samples of absolute pollinator abundance, one could calibrate InVEST's relative indices to predict actual pollinator abundances. And with specific yield functions, one could use these actual abundances to estimate absolute estimates of economic value. This would require, beyond these additional data, custom modeling steps that InVEST does not offer. InVEST does produce, however, the intermediate results necessary to insert these modeling steps. Furthermore, the logic that increasing pollinator abundance and diversity lead to increased yield is supported by previous research (Greenleaf and Kremen 2006).

One option for overcoming this limitation is to link this model with an agricultural production model (InVEST or another), which will take pollinator abundance as one input to predict and map agricultural yields. In formal terms, it will use pollination as a factor in a "production function" that relates yields of a given crop to the quantity and quality of various inputs (e.g., water, soil fertility, labor, chemicals, pollination). Using these production functions, it is possible to estimate the proportion of crop productivity that is due to pollination, and thus the economic value of those pollinators.

Second, the model does not include the dynamics of bee populations over time, and therefore cannot evaluate whether these populations are sustainable given the current landscape. Instead, the model simply provides a static snapshot of the number of pollinators on each cell in the landscape, given simple estimates of nesting sites and food resources. Some of the factors that influence bee populations, like habitat disturbances and typical population fluctuations, are not captured.

Third, the model does not account for the sizes of habitat patches in estimating abundance. For many species, there is a minimum patch size, under which a patch cannot support that species over the long term. There is some evidence that small patches support fewer species of bees (Kremen et al. 2004), but bees can also survive in surprisingly small areas of suitable habitat (Ricketts 2004).

Fourth, pollinators are likely to be influenced by fine-scale features in the landscape, which are difficult to capture in typical land-cover data with typical resolutions of 1km or even 30m. For example, small patches of flower resources in an otherwise hostile habitat for bees can provide important food resources, but will not be detected by typical land cover maps. Some bees are also able to nest in small but suitable areas (a single suitable roadside or tree hollow). Using average values of nesting site or flower availability for each land cover type, along with 30m pixels or larger, will therefore not capture these fine scale but important areas of resources.

Finally, the model does not include managed pollinators, such as honey bees, that are managed in boxed hives and can be moved among fields to pollinate crops. InVEST focuses on the ecosystem service of pollination by bees living wild in the landscape. Managed pollinators are a technological substitute for this ecosystem service, much as a water filtration plant is a substitute for purification services by wetlands and other natural systems. Clearly, any natural resource assessment needs to consider the costs and benefits of investments in technology (filtration plants, managed bees) alongside those of investments into natural capital (wetlands, wild bee pollination).

4.4.4 Data Needs

The model uses five forms of input data (three are required, and two are optional):

1. **Current land cover map (required)**. A GIS raster dataset, with a land use and land cover (LULC) code for each cell. The dataset should be projected in meters and the projection should be defined. This coverage must be of fine enough resolution (i.e., sufficiently small cell-size) to capture the movements of bees on a landscape. If bees fly 800 meters on average and cells are 1000 meters across, the model will not fully capture the movement of bees from their nesting sites to neighboring farms.

   Name: file can be named anything, but avoid spaces (e.g. use lulc_samp_cur)

   Format: standard GIS raster file (e.g., ESRI GRID or IMG), with a column labeled 'value' that designates the LULC class code for each cell (e.g., 1 for forest, 3 for grassland, etc.) The LULC 'value' codes must match LULC class codes used in the Land Attributes table described below. The table can have additional fields, but the only field used in this analysis is one for LULC class code.

   The model also requests three pieces of information about this LULC map, which are optional but will be prompted for in the interface.

   1. The year depicted by the LULC map (optional). You can indicate the year of the LULC map, if known, to designate model runs performed at different time periods (i.e., future scenarios).
   2. The resolution at which the model should run (optional). You can indicate a coarser resolution than that of the native LULC map to prompt the model to resample at this new resolution and to speed up run time. For example, you could run the model at a 200m resolution with a 30m resolution LULC map. If you leave this line blank, the model will perform the analysis at the same resolution of the native LULC map (i.e., the default). (Note: a resolution that is finer than the native resolution of the raster dataset cannot be defined).
   3. Agricultural land cover and land use classes (optional). You can specify LULC classes that represent agricultural parcels dependent upon or that benefit from pollination by bees. Doing so will restrict the calculation of pollinator abundance to only the designated farms. Enter the LULC values in the format 2;9;13;etc. If you do not specify agricultural classes then a farm abundance map will be calculated for the entire landscape (the default). Refer to Klein et al. 2007 for a list of crops and their level of pollinator-dependency.

2. **Table of pollinator species or guilds (required)**. A table containing information on each species or guild of pollinator to be modeled. Guild refers to a group of bee species that show the same nesting behavior, whether preferring to build nests in the ground, in tree cavities, or other habitat features. If multiple species are known to be important pollinators, and if they differ in terms of flight season, nesting requirements, or flight distance, provide data on each separately. If little or no data are available, create a single 'proto-pollinator,' with data taken from average values or expert opinion about the whole pollinator community.

   Name: file can be named anything

   File Type: *.dbf, Excel worksheets (*.xls, .xlsx), or Ms Access tables (*.mdb, .accdb). If using ArcGIS 9.2x then you will need to use .xls or .mdb files. Excel 2007 (.xlsx) and Ms Access 2007 (.accdb) files will only work with ArcGIS 9.3x.

   Rows: each row is a unique species or guild of pollinator.

   Columns: columns contain data on each species or guild. Column order doesn't matter, but columns must be named as follows (italicized portions of names can be customized for meaning, but must be consistent with names in other tables):

   1. *Species*: Name of species or guild (Note: species names can be numerical codes or names. The model will produce outputs coded by the first 4 characters of each species name (e.g., Andr for Andrena nivalis), thus, each species or guild should be uniquely identifiable at 4 characters. If

species or guild are not uniqueluely identifiable at 4 characters then the model will truncate the names at 3 and at a digit).

2. *NS_nest1*, *NS_nest2*, etc.: Nesting guilds of each pollinator. Values should be entered either as 0 or 1, with 1 indicating a nesting type that is utilized and 0 indicating a non-utilized nest type. If a pollinator falls within multiple nesting guilds, then indicate 1s for all compatible nest types. Nesting types might be ground nests, tree cavities, etc.

3. *FS_season1*, *FS_season2*, etc.: Pollinator activity by floral season (i.e., flight season). Values should be entered on a scale of 0 to 1, with 1 indicating the time of highest activity for the guild or species, and 0 indicating no activity. Intermediate proportions indicate the relative seasonal activity. Activity level by a given species over all seasons should sum to 1. Create a different column for each season. Seasons might be spring, summer, fall; wet, dry, etc.

4. *Alpha*: average (or typical) distance each species or guild travels to forage on flowers, specified in meters. InVEST uses this estimated distance to define the neighborhood of available flowers around a given cell, and to weight the sums of floral resources and pollinator abundances on farms. You can determine typical foraging distance of a bee species based on a simple allometric relationship with body size (see Greenleaf et al. 2007).

Example: A hypothetical study with four species. There are two main nesting types, "cavity" and "ground." Species A is exclusively a cavity nester, species B and D are exclusively ground nesters, and species C uses both nest types. There is only a single flowering season, "Allyear," in which all species are active. Typical flight distances, specified in meters (Alpha), vary widely among species.

Species	NS_cavity	NS_ground	FS_allyear	Alpha
A	1	0	1	1490
B	0	1	1	38
C	1	1	1	890
D	0	1	1	84

3. **Table of land cover attributes (required)**. A table containing data on each class in the LULC map (as described above in #1). Data needed are relative indices (0-1), not absolute numbers. Data can be summarized from field surveys, or obtained by expert assessment if field data is unavailable. Name: file can be named anything

   File type: *.dbf, Excel worksheets (*.xls, .xlsx), or Ms Access tables (*.mdb, .accdb). If using ArcGIS 9.2x then you will need to use .xls or .mdb files. Excel 2007 (.xlsx) and Ms Access 2007 (.accdb) files will only work with ArcGIS 9.3x.

   Rows: each row is a different LULC class.

   Columns: each column contains a different attribute of each LULC class, and must be named as follows:

   1. *LULC*: Land use and land cover class code. LULC codes match the 'values' column in the LULC raster and must be numeric, in consecutive order, and unique.

   2. *LULCname*: Descriptive name of LULC class (optional).

   3. *N_nest1*, *N_nest2*, etc.: Relative index of the availability of nesting type 1, 2, etc. within each LULC type, on a scale of 0-1 (values do not need to sum to 1 across nesting types). Set the LULC type with the greatest availability of nesting habitat at 1, and give all other land classes a value in proportion to this maximum value. The italicized parts of names must match those in NS_nest1, etc. in the Table of pollinator species or guilds (described in input #2 above).

   4. *F_season1*, *F_season2*, etc.: Relative abundance (0-1) of flowers in each LULC class for season 1, season 2, etc. There are two aspects to consider when estimate relative floral abundance of each LULC class: % floral abundance or % floral coverage as well as the duration of flowering during each season. For example, a land cover type that comprises 100% of a mass flowering crop that flowers the entire season with an abundance cover of 80% would be given a suitability value of 0.80. A land cover type that flowers only half of the season at 80% floral coverage would be given a floral

suitability value of 0.40. Italicized parts of names must match those in FS_nest1, etc. in the Table of pollinator species or guild file (described in input #2 above).

Example: The same hypothetical study with five LULC classes. Class 1 (Forest) contains the maximum availability of sites for both nesting types ("cavity" and "ground"). The five habitat types vary strongly in flower resources in the single (simplified, year-round) flowering season. Note matching column heads between this table and the Table of pollinator species or guilds.

LULC	LULCname	N_cavity	N_ground	F_allyear
1	Forest	1.0	1.0	1.0
2	Coffee	0.2	0.1	0.5
3	Pasture/grass	0.2	0.1	0.3
4	Shrub/undergrowth	0.2	0.1	0.2
5	Open/urban	0.2	0.1	0.3

In this case the agricultural land-use, coffee, is perennial and has some cavity and ground nesting resources. In a more frequently disturbed annual cropping system, nesting resources may be 0. For large monoculture cropping systems, floral resources are only available during a single crop's blooming period, which may be as brief a period as a few weeks, and therefore not provide a very reliable resource for pollinators. It is important to consider carefully what the cropping system of interest realistically provides in the way of floral and nesting resources, because overestimating the value of cropland as a resource to pollinators will underestimate the value of natural habitat to pollinators. If different crop fields have different cropping systems and therefore different relative magnitudes of pollinator resources, it would be best to reclassify the land-use map to create a different land-use class for each cropping system.

4. **Half-saturation constant (optional)**. The model will also prompt you to enter a half-saturation constant, which will be used when calculating the pollinator service value map. This constant converts the pollinator supply into yield and represents the abundance of pollinators required to reach 50% of pollinator-dependent yield. We suggest that the user apply the default value derived from previous work (i.e., 0.125, Lonsdorf et al 2009) unless there are data to justify changing it. The value must be greater than 0 and it is unlikely that the value would be greater than 0.2.

5. **Future Scenarios (optional)**. To evaluate change in pollination services under a future scenario, a Future Land Cover Map needs to be provided for that future time point (along with the year depicted). The raster dataset needs to be formatted exactly like the current Land Cover Map (data input #1). This LULC map could reflect changes in land management policy, trends in land use change (e.g., agricultural expansion, urbanization, increased habitat protection).

4.4.5 Running the Model

The model is available as a standalone application accessible from the Windows start menu. For Windows 7 or earlier, this can be found under *All Programs -> InVEST +VERSION+ -> Pollination*. Windows 8 users can find the application by pressing the windows start key and typing "pollination" to refine the list of applications. The standalone can also be found directly in the InVEST install directory under the subdirectory *invest-3_x86/invest_pollination.exe*.

Viewing Output from the Model

Upon successful completion of the model, a file explorer window will open to the output workspace specified in the model run. This directory contains an *output* folder holding files generated by this model. Those files can be viewed in any GIS tool such as ArcGIS, or QGIS. These files are described below in Section *Interpreting Results*.

4.4.6 Interpreting Results

Final Results

Final results are found in the *Output* folder within the *Workspace* specified for this module.

- **Parameter log**: Each time the model is run, a text (.txt) file will appear in the *Output* folder. The file will list the parameter values for that run and will be named according to the service, the date and time, and the suffix.
- **sup_tot_cur**: This is a map of pollinator abundance index, summing over all bee species or guilds. It represents an index of the likely abundance of pollinator species nesting on each cell in the landscape, given the availability of nesting sites and of flower (food) resources nearby.
- **sup_tot_fut**: The same as above, but for the future scenario land cover map, if provided.
- **frm_avg_cur**: This is a map of pollinator abundance on each agricultural cell in the landscape, based on the average of all bee species or guilds. It represents the likely average abundance of pollinators visiting each farm site.
- **frm_avg_fut**: The same as above, but for the future scenario land cover map, if provided.
- **sup_val_cur**: This is a map of "pollinator service value": the relative value of the pollinator "supply" in each agricultural cell to crop production in the surrounding neighborhood. It is an index derived by distributing the values in *frm_val_cur* (an intermediate result) back to surrounding pollinator sources, using information on flight ranges of contributing pollinators. This is a map of where pollination services are coming from, and their (relative) values. Units are not dollars per se, but the index is a relative measure of economic value.
- **sup_val_fut**: The same as above, but for future scenario land cover map, if provided.

Intermediate Results

You may also want to examine the intermediate results. These files can help determine the reasons for the patterns in the final results. They are found in the *Intermediate* folder within the *Workspace* specified for this module.

- **hn_<beename>_cur**: This is a map of the availability of nesting sites for each pollinator. The map depends on the values you provide for the availability of each nesting type in each LULC class, and for the nesting habits of each bee species. In fact, values in this map are simply the product of those two provided numbers (e.g., in the example tables given above, species A is entirely a cavity nester, and coffee has a 0.2 value for cavity nest availability, so the value for species A in a coffee cell will be 1 x 0.2 = 0.2). (Note: the "<beename>" portion of each file name will be the first 4 characters of the 'Species' column in dataset #2, so make sure these 4 characters identify each species or guild uniquely).
- **hn_<beename>_fut**: The same as above, but for the future scenario land cover map, if provided.
- **hf_<beename>_cur**: This is a map of availability of flower resources for each species in the neighborhood around each cell. The value for each cell is a sum of surrounding flower values, with values from nearer cells given more weight than those from cells further away. The sum is taken over a neighborhood with the radius equal to the typical flight range of the bee (i.e., 'Alpha' in dataset #2).
- **hf_<beename>_fut**: The same as above, but for the future scenario land cover map, if provided.
- **sup_<beename>_cur**: This is a map of the pollinator abundance index for each bee species or guild modeled. There will be a different map for each species or guild included in your analysis. This map represents the relative likely abundance of a pollinator species nesting on each cell in the landscape, given the availability of nesting sites there and of flower (food) resources nearby.
- **sup_<beename>_fut**: The same as above, but for the future scenario land cover map, if provided.

- **frm_<beename>_cur**: This is a map of the abundance index for each bee species or guild on each agricultural cell in the landscape. There will be a different map for each species or guild included in your analysis. If you did not specify agricultural classes, then every cell (and land cover classes) in the LULC map will contain values.
- **frm_<beename>_fut**: The same as above, but for the future scenario land cover map, if provided.
- **frm_val_cur**: This is a map of "farm value": the relative value of crop production on each agricultural cell due to wild pollinators. It is based on a transformation of *frm_ave_cur*, using a simple saturating yield function to translate abundance units into value units. It represents, in terms of crop production, the contribution of wild pollinators. Units are not dollars per se, but the index is a relative measure of economic value.
- **frm_val_fut**: The same as above, but for future scenario land cover map, if provided.

4.4.7 Appendix: Data Sources

List of globally important crops and their dependence on animal pollinators: (Klein et al. 2007).

4.4.8 References

Allen-Wardell, G., P. Bernhardt, R. Bitner, A. Burquez, S. Buchmann, J. Cane, PA Cox, V. Dalton, P. Feinsinger, M. Ingram, D. Inouye, CE Jones, K. Kennedy, P. Kevan, and H. Koopowitz. 1998. The potential consequences of pollinator declines on the conservation of biodiversity and stability of food crop yields. Conservation Biology 12: 8-17.

Cane, JH. 1997. Lifetime monetary value of individual pollinators: the bee habropoda laboriosa at rabbiteye blueberry (vaccinium ashei reade). Acta Horticulturae 446: 67-70.

Costanza, R., R. d'Arge, R. de Groot, S. Farber, M. Grasso, B. Hannon, K. Limburg, S. Naeem, RV O'Neill, J. Paruelo, RG Raskin, P. Sutton, and M. van den Belt. 1997. The value of the world's ecosystem services and natural capital. Nature 387: 253-260.

Free, JB. 1993. Insect pollination of crops. Academic Press, London.

Greenleaf, SS, NM Williams, R. Winfree, and C. Kremen. 2007. Bee foraging ranges and their relationship to body size. Oecologia 153: 589-596.

Greenleaf, SS, and C. Kremen. 2006. Wild bee species increase tomato production and respond differently to surrounding land use in Northern California. Biological Conservation 133:81-87.

Klein, AM, BE Vaissiere, JH Cane, I. Steffan-Dewenter, SA Cunningham, C. Kremen, and T. Tscharntke. 2007. Importance of pollinators in changing landscapes for world crops. Proceedings of the Royal Society B-Biological Sciences 274: 303-313.

Kremen, C., NM Williams, RL Bugg, JP Fay, and RW Thorp. 2004. The area requirements of an ecosystem service: crop pollination by native bee communities in California. Ecology Letters 7: 1109-1119.

Lonsdorf, E., C. Kremen, T. Ricketts, R. Winfree, N. Williams, and SS Greenleaf. 2009. Modelling pollination services across agricultural landscapes. Annals of Botany 1: 12 online [http://aob.oxfordjournals.org/content/103/9/1589.abstract?sid=b0f6fc9e-bd52-4bec-8248-50e56b78344d]

Lonsdorf, E., TH Ricketts, CM Kremen, NM Williams, and S. Greenleaf. in press. Pollination services in P. Kareiva, TH Ricketts, GC Daily, H. Tallis, and S. Polasky, eds. The theory and practice of ecosystem service valuation.

Losey, JE, and M. Vaughan. 2006. The economic value of ecological services provided by insects. Bioscience 56: 311-323.

Nabhan, GP, and SL Buchmann. 1997. Services provided by pollinators. Pages 133-150 in GC Daily, ed. Nature's services. Island Press, Washington, D.C.

Priess, JA, M. Mimler, AM Klein, S. Schwarze, T. Tscharntke, and I. Steffan-Dewenter. 2007. Linking deforestation scenarios to pollination services and economic returns in coffee agroforestry systems. Ecological Applications 17: 407-417.

Ricketts, TH. 2004. Tropical forest fragments enhance pollinator activity in nearby coffee crops. Conservation Biology 18: 1262-1271.

Ricketts, TH, NM Williams, and MM Mayfield. 2006. Connectivity and ecosystem services: crop pollination in agricultural landscapes. Pages 255-289 in M. Sanjayan and K. Crooks, eds. Connectivity for Conservation. Cambridge University Press, Cambridge, UK.

Roubik, DW, and M. Aluja. 1983. Flight ranges of Melipona and Trigona in tropical forest. Journal of the Kansas Entomological Society 56: 217-222.

Southwick, EE, and L. Southwick. 1992. Estimating the economic value of honey-bees (Hymenoptera; Apidae) as agricultural pollinators in the United States. Journal of Economic Entomology 85: 621-633.

Winfree, R., J. Dushoff, EE Crone, CB Schultz, RV Budny, NM Williams, and C. Kremen. 2005. Testing simple indices of habitat proximity. American Naturalist 165(6): 707-717.

CHAPTER

FIVE

FINAL ECOSYSTEM SERVICES:

5.1 Forest Carbon Edge Effect Model

5.1.1 Summary

The InVEST carbon edge model extends the approach of the InVEST carbon model to account for forest carbon stock degradation due to the creation of forest edges. It applies known relationships between carbon storage and distance from forest edge to calculate edge effects in carbon storage, and combines these estimates with carbon inventory data to construct the overall carbon map. The model for edge effects pertains to above-ground carbon only, because edge effects have not been documented for the other carbon pools. For all other carbon pools, and for non-tropical forest classes, or if the model is run without edge effects, it follows the IPCC (2006) inventory approach to assigning carbon storage values by land cover class.

5.1.2 Introduction

Effects of future land-use change on carbon storage or sequestration can be modeled by applying carbon storage estimates found in the literature for different habitat types to each habitat found in a landscape (e.g., Nelson et al. 2010). However, this approach assumes that all habitat is equivalent in its quality of carbon storage, regardless of where it occurs, despite the fact that there is substantial evidence that fragmentation can play a dramatic role in altering carbon storage and sequestration rates in tropical forests (Broadbent et al. 2008, Dantas et al 2011, Laurance et al. 1997, 2000, 2001, 2002). For example, core forest has been shown to store more than three times the carbon of edge forest in Brazilian Atlantic forests (Dantas et al. 2011). Chaplin-Kramer et al. (REF) investigated this pattern for the entire pantropics using remotely sensed data on biomass (Baccini et al 2012) and associated land cover (Friedl et al. 2011) and found a continuous relationship of carbon storage degrading gradually toward a forest edge, which varies substantially from region to region. This model accounts for these documented edge effects in tropical forests, assigning carbon storage based on the distance of a pixel to the nearest forest edge. This can improve the overall accuracy of carbon storage estimates by 20% relative to forest inventory approaches, and better captures the impacts of forest degradation from fragmentation, beyond merely the area of habitat lost.

5.1.3 The model

The InVEST carbon edge model is an update to the InVEST carbon model, which incorporates the degradation of carbon storage that occurs due to edge effects in tropical forests. The user designates which land cover classes are forest, and then the model uses pregenerated regression results to predict the carbon throughout a forest parcel based on its distance to the nearest forest edge. These results are combined with the carbon assigned to non-forest classes through traditional inventory methods (IPCC 2006) used in the InVEST carbon model, to generate a map of above-ground carbon storage for all land cover classes. The InVEST carbon edge effects model can be run to calculate only above ground carbon or all carbon pools, and it can be run with or without edge effects. It is important to note that the edge effects regression only pertains to above-ground carbon stocks because edge effects have only been detected for above-ground biomass. To include the other three carbon pools (below-ground, soil, and standing dead matter), carbon density (Mg/ha) should be included for each land cover class in the biophysical table.

How it works

This model follows the methodology described in Chaplin-Kramer et al. (REF), which constructs a series of regression models between forest biomass density (Mg/ha) and distance from forest edge (km) for 100 km x 100 km grid cells throughout the pantropics. In grid cells where the majority of pixels were from forest biomes, three candidate regression models are considered to represent the relationship between biomass density and distance to forest edge:

1. Asymptotic: Biomass $= \theta_1 - \theta_2 \cdot \exp(-\theta_3 \cdot \text{Distance})$
2. Logarithmic: Biomass $= \theta_1 + \theta_2 \cdot ln(\text{Distance})$
3. Linear: Biomass $= \theta_1 + \theta_2 \cdot \text{Distance}$

Then, for each grid cell, the candidate with the highest r^2 value is used to best represent the relationship between density and distance to forest edge. Models (2) and (3) were deemed as suitable (and more simplistic) alternatives in cells where higher distances were generally not observed and as a result the forest core was not firmly established. In the vast majority of grid cells, model (1) was optimal.

The results of these regressions can be found in the carbon edge regression parameter shapefile (forest_carbon_edge_regression_model_parameters.shp) in the core dataset that ships with the InVEST carbon edge model. For any forest pixel within the study region, the model calculates the distance of that pixel from forest edge then calculates biomass to a predefined number of nearest regression models which is then aggregated to a single result using a distance linear interpolation scheme. The model then converts biomass to carbon with a user provided conversion factor, defaulted to 0.47 (IPCC 2006). The user can designate the number of local models used in the interpolation scheme which is defaulted to 10 but can range anywhere from 1 (only closest point) to 2635 (every regression model on the planet). Note that a selection of 1 may result in artificially large differences in carbon when moving from one pixel to the next where they fall in different regression grid cells. The higher the number of regression grid cells selected, the smoother the transition from one pixel to the next. The user may wish to select the number of grid cells overlapping the entire study region in order to eliminate any artifacts of model selection. This can be determined by examining the intermediate_outputs\local_carbon_shape.shp geometry overlaid on the area of interest. The linear interpolation scheme for biomass b on pixel p is given below

$$b_p = \frac{\sum_{i\in n} \frac{1}{d_i} b_i'}{\sum \frac{1}{d_i}}$$

Where,

- $b_p$is the interpolated biomass on pixel p
- n is the number of nearest models to interpolate from, a value provided by the user
- i is the i^{th} nearest biomass model from pixel p
- d_i is the distance from pixel p to the centroid of the i^{th} biomass model.

The carbon calculated for non-forest classes follows the methodology from the InVEST carbon model, assigning values based on forest carbon inventory data designated in the biophysical table. The carbon maps following this inventory approach and the edge effects approach are merged into the final carbon map, such that the forest land covers exhibit edge effects and all other land covers will not.

5.1.4 Data needs

The model uses 6 forms of input data. Five are required and one is optional.

1. Land-use/cover map (required). A map of the different types of land uses or land cover (LULC) for the study region, classified into different categories.

   Name: file can be named anything (carbon_edge_lulc_demo.tif in the sample data)

   Format: standard GIS raster file (e.g., ESRI GRID or IMG), with a column labeled 'value' that designates the LULC class code for each cell (integers only; e.g., 1 for forest, 10 for grassland, etc.)

1. Biophysical table (required). A table providing information about which classes in the land-use/land-cover map (input 5) are considered forest and should have the edge effect regression applied, and carbon density (Mg per ha) for the land cover classes that are not forest. If "all pools" is selected for "carbon pools to calculate", columns iii-vi must be included (though 0's can be placed for any pools that you do not wish to calculate); if "above ground only" is selected, columns iv-vi can be excluded.

   Name: file can be named anything (edge_carbon_lu_table.csv in sample data)

   Type: *.csv

   Columns: the columns must be named as follows:

   (a) lucode

   (b) is_tropical_forest

   (c) c_above

   (d) c_below

   (e) c_soil

   (f) c_dead

   (g) Description

   Example (data for Brazil):

lucode	Is_tropical_forest	c_above	c_below	c_soil	c_dead	Description
0	0	0	0	0	0	Water
1	1	n/a	16.8	60	14.4	Evergreen Needleleaf Forest
2	1	n/a	22.4	60	10.2	Evergreen Broadleaf Forest
3	1	n/a	12.8	60	11	Deciduous Needleleaf Forest
4	1	n/a	15.6	60	13.4	Deciduous Broadleaf Forest
5	1	n/a	14.1	60	12.1	Mixed Forest
6	0	10.5[1]	6.7	60.1	1.3	Closed Shrublands
7	0	4.5[1]	7.3	65.5	1.1	Open Shrublands
8	0	20.1[2]	23.6	151.4	2.6	Woody Savannas
9	0	4.4[2]	7.1	211	1.5	Savanna
10	0	3[1]	7.9	71	0.9	Grasslands
12	0	0.25[3]	0.5	50.8	2.4	Croplands
13	0	0	0	0	0	Urban/Built-up
16	0	0	0	0	0	Barren or sparsely vegetated

Source: (1) Miranda, Sabrina do Couto, et al. Regional variations in biomass distribution in Brazilian savanna woodland. Biotropica 46.2 (2014): 125-138. (2) Saatchi, Susan S., et al. Distribution of aboveground live biomass in the Amazon basin. Global Change Biology 13.4 (2007): 816-837. (3) Fearnside, Philip M. Greenhouse gases from deforestation in Brazilian Amazonia: net committed emissions. Climatic Change 35.3 (1997): 321-360.

1. Carbon pools to calculate (required, select from menu): user must select either "all carbon pools" or "above ground only".

2. Compute forest edge (optional): check this box if edge effects on above-ground biomass should be included. If this box is checked inputs 5-7 are required.

3. Carbon edge regression parameter (optional, included in base data): this dataset was derived from a pantropical analysis of the relationship between forest biomass and distance from forest edge (Chaplin-Kramer et al. REF). The rows contain the coefficients of the unique parameters for each 100 x 100 km subregion across the tropics. This information need not be altered unless you have run a separate regression for your region and have better or updated information.

   Name: carbon_edge_regression_model_parameters, in the "core data" folder included in the carbon edge model installer

   Format: shapefile

4. Number of nearest model point to average (optional): used when calculating the biomass in a pixel. This number determines the number of closest regression models that are used when calculating the total biomass. Each local model is linearly weighted by distance such that the biomass in the pixel is a function of each of these points with the closest point having the highest effect. Default value is 10. Higher values smooth the variation in the edge effect detected in the different grid cells (seen in the carbon edge regression parameter shapefile, input 3) to a greater degree.

5. Carbon conversion factor (optional): enter the number by which to multiply forest above-ground biomass to convert to carbon. Default value is 0.47 (according to IPCC 2006). This pertains to the edge-effects regression parameters only; all values in the biophysical table should already be in terms of carbon, not biomass.

6. Service areas of interest (optional): If a summary of the carbon value is desired, a vector dataset containing the area(s) of interest, either as a region area or partitioned into subregions (e.g., ecoregions, districts, etc.) can be included.

   Name: file can be named anything (demo_servicesheds in the sample data)

   Format: a vector (polygon) file

5.1.5 Running the model

The model is available as a standalone application accessible from the install directory of InVEST (under the subdirectory invest-3_x86/invest_carbon_edge_effect.exe).

Viewing Output from the Model

Upon successful completion of the model, a file explorer window will open to the output workspace specified in the model run. This directory contains an output folder holding files generated by this model. Those files can be viewed in any GIS tool such as ArcGIS, or QGIS. These files are described below in Section Interpreting Results.

5.1.6 Interpreting Results

Final Results

Final results are found in the *Workspace* folder specified for this module.

- **carbon_edge_effect-log: Each time the model is run, a text** (.txt) file will appear in the *Output* folder. The file will list the parameter values for that run and will be named according to the service, the date and time, and the suffix.

- **carbon_map_<suffix>: a map of carbon stock per pixel, with the** amount in forest derived from the regression based on distance to forest edge, and the amount in non-forest classes according to the biophysical table. Note that because the map displays carbon (Mg) per pixel, coarser resolution maps should have higher values for carbon, because the pixel areas are larger.

- **aggregated_carbon_stocks: If an AOI is provided by the user,** this summarizes the total carbon and mean carbon per ha in the areas of interest defined in that AOI.

Intermediate Results

You may also want to examine the intermediate results. These files can help determine the reasons for the patterns in the final results. They are found in the intermediate_outputs folder within the *Workspace* specified for this module.

- **intermediate_outputs\local_carbon_shape: this is the** regression parameters reprojected to match your study area
- **intermediate_outputs\edge_distance_<suffix>: the distance** of each forest pixel to the nearest forest edge
- **intermediate_outputs\forest_edge_carbon_stocks_<suffix>:** a map of carbon in the forest only, according to the regression method
- **intermediate_outputs\non_forest_carbon_stocks_<suffix>: a** map of carbon in the non-forest classes only, according to the carbon inventory listed in the biophysical table

5.1.7 References

Baccini, A., S. J. Goetz, W. S. Walker, N. T. Laporte, M. Sun, D. Sulla-Menashe, J. Hackler, P. S. A. Beck, R. Dubayah, M. A. Friedl, S. Samanta, and R. A. Houghton. 2012. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. Nature Climate Change 2:182–185.

Chaplin-Kramer, R., I. Ramler, R. Sharp, N. M. Haddad, J. S. Gerber, P. C. West, L. Mandle, P. Engstrom, A. Baccini, S. Sim, C. Mueller, and H. King. (n.d.). Degradation in carbon stocks near tropical forest edges. Nature Communications.

Dantas de Paula, M., Alves-Costa, C., Tabarelli, M., 2011. Carbon storage in a fragmented landscape of Atlantic forest: the role played by edge-affected habitats and emergent trees. Tropical Conservation Science 4, 349–358.

Friedl, M. A., D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang. 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. Remote Sensing of Environment 114:168–182.

Intergovernmental Panel on Climate Change (IPCC). 2006. IPCC Guidelines for National Greenhouse Gas Inventories. Volume 4: Agriculture, Forestry and Other Land Use.

Laurance, W. F., 1997. Biomass Collapse in Amazonian Forest Fragments. Science 278, 1117–1118.

Laurance, W.F., 2000. Do edge effects occur over large spatial scales? Trends in ecology & evolution 15, 134–135.

Laurance, William F., Williamson, G.B., 2001. Positive Feedbacks among Forest Fragmentation, Drought, and Climate Change in the Amazon. Conservation Biology 15, 1529–1535.

Laurance, W., Lovejoy, T., Vasconcelos, H., Bruna, E., Didham, R., Stouffer, P., Gascon, C., Bierregaard, R., Laurance, S., Sampaio, E., 2002. Ecosystem decay of Amazonian forest fragments: a 22-year investigation. Conservation Biology 16, 605–618.

Nelson, E., et al. 2010. Projecting global land-use change and its effect on ecosystem service provision and biodiversity with simple models. PLOS One 5: e14327

5.2 Carbon Storage and Sequestration: Climate Regulation

5.2.1 Summary

Terrestrial ecosystems, which store more carbon than the atmosphere, are vital to influencing carbon dioxide-driven climate change. The InVEST model uses maps of land use and land cover types and data on wood harvest rates, harvested product degradation rates, and stocks in four carbon pools (aboveground biomass, belowground biomass, soil, dead organic matter) to estimate the amount of carbon currently stored in a landscape or the amount of carbon sequestered over time. Additional data on the market or social value of sequestered carbon and its annual rate of change, and a discount rate can be used in an optional model that estimates the value of this ecosystem service to society. Limitations of the model include an oversimplified carbon cycle, an assumed linear change in carbon sequestration over time, and potentially inaccurate discounting rates.

5.2.2 Introduction

Ecosystems regulate Earth's climate by adding and removing greenhouse gases (GHG) such as CO_2 from the atmosphere. In fact, forests, grasslands, peat swamps, and other terrestrial ecosystems collectively store much more carbon than does the atmosphere (Lal 2002). By storing this carbon in wood, other biomass, and soil, ecosystems keep CO_2 out of the atmosphere, where it would contribute to climate change. Beyond just storing carbon, many systems also continue to accumulate it in plants and soil over time, thereby "sequestering" additional carbon each year. Disturbing these systems with fire, disease, or vegetation conversion (e.g., land use / land cover (LULC) conversion) can release large amounts of CO_2. Other management changes, like forest restoration or alternative agricultural practices, can lead to the storage of large amounts of CO_2. Therefore, the ways in which we manage terrestrial ecosystems are critical to regulating our climate.

As with all other models for which InVEST provides estimates of value, we are focused on the social value of carbon sequestration and storage. Terrestrial-based carbon sequestration and storage is perhaps the most widely recognized of all ecosystem services (Stern 2007, IPCC 2006, Canadell and Raupach 2008, Capoor and Ambrosi 2008, Hamilton et al. 2008, Pagiola 2008). The social value of a sequestered ton of carbon is equal to the social damage avoided by not releasing the ton of carbon into the atmosphere (Tol 2005, Stern 2007). Calculations of social cost are complicated and controversial (see Weitzman 2007 and Nordhaus 2007b), but have resulted in value estimates that range from USD \$9.55 to \$84.55 per metric ton of CO_2 released into the atmosphere (Nordhaus 2007a and Stern 2007, respectively).

In addition to the social value of carbon sequestration and storage, there are several emerging markets for carbon based on both regulation and voluntary demand. The Kyoto Protocol – the current treaty addressing international climate change – includes a mechanism for establishing projects that sequester carbon to earn credits, which they then can sell to others needing to offset their own CO_2 emissions. As a result of the Kyoto Protocol, the European Union Emissions Trading Scheme (EU ETS) emerged to allow the regulated firms of the EU to trade their emissions allowances. The Chicago Climate Exchange (CCX) emerged in the United States, which is not a signatory party of the Kyoto Protocol. The CCX allows interested parties to trade emissions offsets that have been certified on a voluntary basis. The EU ETS and the CCX hadprices of around 25 Euros and \$USD 6 per metric ton of CO_2, respectively as of April 2008. In addition to these centralized markets, there is a substantial over-the-counter market for voluntary carbon offsets. For details about the price of these offsets, see Conte and Kotchen (2010).

Currently these markets only apply to carbon sequestration (i.e., the additional storage of carbon over time), but there is increased interest in financial incentives to avoid release of carbon from ecosystems in the first place, so-called "reduced emissions from deforestation and degradation" or "REDD" (Gibbs et al. 2007, Mollicone et al. 2007, Mackey et al. 2008). This option was accepted during the last meeting of the parties to the UN Framework Convention on Climate Change and is likely to be written in to the follow up agreement to the Kyoto Protocol. Payments for REDD would financially reward forest owners for reversing their planned deforesting and thinning actions (Sedjo and Sohngen 2007, Sohngen et al. 2008). Issues of accounting and verification have slowed the emergence of REDD markets, but many are anticipating them with private transactions.

While market prices are one way to estimate the value of CO_2 sequestration, these prices will reflect policies, subsidies, and other factors, and therefore will only indicate the true value of this service to society by chance (Murray et al. 2007). For this reason, we recommend that users rely on the avoided damages associated with the emission of CO_2 into the atmosphere rather than prices in existing carbon markets to estimate the social value of carbon sequestration and storage.

Managing landscapes for carbon storage and sequestration requires information about how much and where carbon is stored, how much carbon is sequestered or lost over time, and how shifts in land use affect the amount of carbon stored and sequestered over time. Since land managers must choose among sites for protection, harvest, or development, maps of carbon storage and sequestration are ideal for supporting decisions influencing these ecosystem services.

Such maps can support a range of decisions by governments, NGOs, and businesses. For example, governments can use them to identify opportunities to earn credits for reduced (carbon) emissions from deforestation and degradation (REDD). Knowing which parts of a landscape store the most carbon would help governments efficiently target incentives to landowners in exchange for forest conservation. Additionally, a conservation NGO may wish to invest in areas where high levels of biodiversity and carbon sequestration overlap (Nelson et al. 2008). A timber company may also want to maximize its returns from both timber production and REDD carbon credits (Plantinga and Birdsey 1994), in which case they could use the InVEST timber production model in tandem with the carbon model to assess management options.

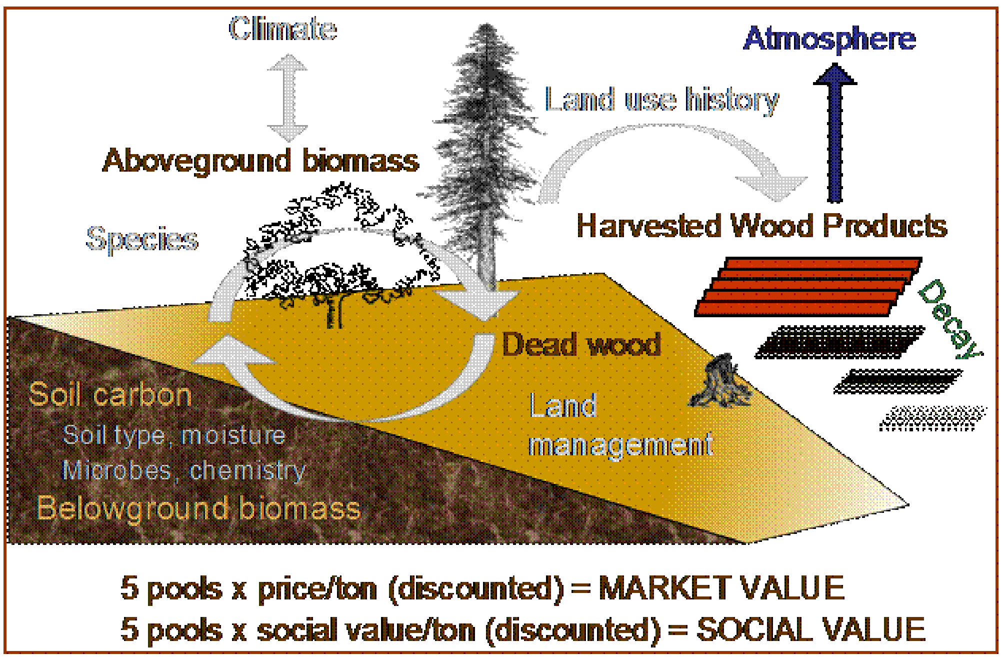


Figure 1. Conceptual model of carbon storage and sequestration. Parameters depicted in color are included in the InVEST model, while those in gray are not.

5.2.3 The Model

Carbon storage on a land parcel largely depends on the sizes of four carbon "pools:" aboveground biomass, belowground biomass, soil, and dead organic matter (Fig. 1). The InVEST Carbon Storage and Sequestration model aggregates the amount of carbon stored in these pools according to the land use maps and classifications produced by the user. Aboveground biomass comprises all living plant material above the soil (e.g., bark, trunks, branches, leaves). Belowground biomass encompasses the living root systems of aboveground biomass.Soil organic matter is the organic component of soil, and represents the largest terrestrial carbon pool. Dead organic matter includes litter as well as lying and standing dead wood. A fifth optional pool included in the model applies to parcels that produce harvested wood products (HWPs) such as firewood or charcoal or more long-lived products such as house timbers or furniture. Tracking carbon in this pool is useful because it represents the amount of carbon kept from the atmosphere by a given product.

Using maps of land use and land cover types and the amount of carbon stored in carbon pools, this model estimates: the net amount of carbon stored in a land parcel over time; the total biomass removed from a harvested area of the parcel, and the market and social values of the carbon sequestered in remaining stock. Limitations of the model include an oversimplified carbon cycle, an assumed linear change in carbon sequestration over time, and potentially inaccurate discounting rates. Biophysical conditions important for carbon sequestration such as photosynthesis rates and the presence of active soil organisms are also not included in the model (Fig. 1).

How it Works

The model runs on a gridded map of cells called raster format in GIS. If the HWP pool is included in the analysis, a polygon map of harvest parcels is also modeled. Each cell in the raster is assigned a land use and land use and land cover (LULC) type such as forest, pasture, or agricultural land. Each harvest polygon is assigned harvest type referring to the harvested product, harvest frequency, and product decay rates. After running the model in raster format, results can be summarized to practical land units such as individual properties, political units, or watersheds.

For each LULC type, the model requires an estimate of the amount of carbon in at least one of the four fundamental pools described above. If the user has data for more than one pool, the modeled results will be more complete. The model simply applies these estimates to the LULC map to produce a map of carbon storage in the carbon pools included.

For the fifth carbon pool, HWP, model values are defined for each parcel (polygon) and not for each LULC. For each parcel the user indicates the amount of biomass, in terms of carbon, removed per harvest, the frequency of harvests, and the rate at which the products that contain carbon degrade. With these data, the model calculates the amount of stored carbon that originated in a parcel but now resides in finished products such as houses or furniture. The model converts parcel level HWP carbon values into a grid cell layer that spatially matches the grid system used for the other four carbon storage pools.

The model aggregates the carbon in each of the five pools, providing an estimate of total carbon storage in each grid cell and across the whole landscape. If carbon storage data for a given pool are not mapped, then total carbon storage will be underestimated. The model also outputs the total biomass and volume of wood removed from each harvested parcel up to the year associated with the modeled landscape.

If the user provides both a current and future LULC map, then the net change in carbon storage over time (sequestration and loss) and its social value can be calculated. To estimate this change in carbon sequestration over time, the model is simply applied to the current landscape and a projected future landscape, and the difference in storage is calculated, map unit by map unit. If multiple future scenarios are available, the differences between the current and each alternate future landscape can be compared.

Outputs of the model are expressed as Mg of carbon per grid cell, or if desired, the value of sequestration in dollars per grid cell. We strongly recommend using the social value of carbon sequestration if the user is interested in expressing sequestration in monetary units. The social value of a sequestered ton of carbon is the social damage avoided by not releasing the ton of carbon into the atmosphere. The market value may be applicable if the user is interested in identifying the value of the landscape for trading under current market conditions. The market value of terrestrial-based carbon sequestration is the price per metric ton of carbon traded in marketplaces such as the Chicago Climate Exchange (ECX).

The valuation model estimates the economic value of sequestration (not storage) as a function of the amount of carbon sequestered, the monetary value of each unit of carbon, a monetary discount rate, and the change in the value of carbon sequestration over time (Fig. 1). **Thus, valuation can only be done in the carbon model if you have a future scenario.** Valuation is applied to sequestration, not storage, because current market prices relate only to carbon sequestration. Discount rates are multipliers that typically reduce the value of carbon sequestration over time. The first type of discounting, the standard economic procedure of financial discounting, reflects the fact that people typically value immediate benefits more than future benefits due to impatience and uncertain economic growth. The second discount rate adjusts the social value of carbon sequestration over time. This value will change as the impact of carbon emissions on expected climate change-related damages changes. If we expect carbon sequestered today to have a greater impact on climate change mitigation than carbon sequestered in the future this second discount rate should be positive. On the other hand, if we expect carbon sequestered today to have less of an impact on climate change mitigation than carbon sequestered in the future this second discount rate should be negative.

Uncertainty Analysis

In many cases, limited data can make it difficult to determine precisely the amount of carbon in different pools. To accomodate such data limitations, the model optionally performs uncertainty analysis. If users choose to run the model with uncertainty analysis, then inputs and outputs are both affected.

Input data when using uncertainty analysis must specify probability distributions for amount of carbon in different pools. For each carbon pool type, input data must specify both the mean estimate, which represents the expected carbon amount, and the standard deviation, which represents the uncertainty for the estimate.

When running uncertainty analysis, model outputs include all of the original outputs of the non-uncertainty model, such as a raster mapping total carbon per grid cell. To calculate the outputs produced by the non-uncertainty model, the uncertainty model uses the user-provided mean estimates for the carbon pools and ignores the standard deviation data.

In addition to these outputs, which use only the mean estimate data, the uncertainty model also produces two types of uncertainty outputs: (1) 'confidence' rasters to indicate areas where we are confident that sequestration or emissions will occur, and (2) standard deviations for outputs.

Confidence Raster

When provided with uncertainty data, the carbon model will produce a 'confidence' output raster, which uses both the mean and the standard deviation data and highlights areas where it is highly likely that storage will either increase or decrease. The model uses a user-provided confidence threshold as the minimum probability for which grid cells should be highlighted.

To compute the probability that storage increases or decreases in a particular grid cell, we use the LULC data and the HWP data (if present) to construct probability distributions for the current carbon storage in the grid cell and the future carbon storage in the cell. The current carbon storage is distributed with mean μ_{curr} and standard deviation σ_{curr}. The future carbon storage is distributed with mean μ_{fut} and standard deviation σ_{fut}. Since we assume that both are normally distributed, we can compute the probability p that future carbon storage is greater than current carbon storage as follows:

$$p = \Phi(\frac{\mu_{fut} - \mu_{cur}}{\sqrt{\sigma_{curr}^2 + \sigma_{fut}^2}})$$

where Φ is the cumulative distribution function of the normal distribution.

This value of p for a particular grid cell is then used to determine how confident we are that storage will either increase or decrease in that cell.

Output Standard Deviations

In addition to the confidence maps, the uncertainty model will also compute standard deviations for output quantities such as carbon storage, carbon sequestration, and value of sequestered carbon.

These standard deviations are computed via a Monte Carlo simulation. For each iteration of the simulation, the model samples a value for carbon per grid cell for each LULC type, given the input normal distribution for that LULC type. Then, for the given iteration of the simulation, the model will assign that amount of carbon to each pixel with the given LULC type, and compute the amount of carbon stored in each scenario. In other words, for a given run of the iteration, all pixels with the same LULC type will be assigned the same amount of carbon; that amount will be chosen by taking a random sample from the input normal distribution.

Outputs for each run of the simulation are recorded, and then analyzed to extract data about mean and standard deviation, which are reported in the output summary file.

This feature is not supported by the current model if HWP analysis is enabled.

REDD Scenario Analysis

The carbon model can optionally perform scenario analysis according to a framework of Reducing Emissions from Forest Degradation and Deforestation (REDD) or REDD+. REDD is a scheme for emissions reductions under which countries that reduce emissions from deforestation can be financially compensated. REDD+ builds on the original REDD framework by also incorporating conservation, sustainable forest management, and enhancement of existing carbon stocks.

To perform REDD scenario analysis, the model requires three LULC maps: one for the current scenario, one for a future baseline scenario, and one for a future scenario under a REDD policy. The future baseline scenario is used to compute a reference level of emissions against which the REDD scenario can be compared. Depending on the specifics on the desired REDD framework, the baseline scenario can be generated in a number of different ways; for instance, it can be based on historical rates of deforestation or on projections. The REDD policy scenario map reflects future LULC under a REDD policy to prevent deforestation and enhance carbon sequestration.

Based on these three LULC maps for current, baseline, and REDD policy scenarios, the carbon biophysical model produces a number of outputs. First, it produce rasters for total carbon storage for each of the three LULC maps. Second, it produces two sequestration rasters. One sequestration raster indicates sequestration from the current scenario to the baseline scenario. The other sequestration raster indicates sequestration from the current scenario to the REDD policy scenario.

If uncertainty analysis is enabled, the carbon biophysical model will also produce two additional confidence rasters. One raster represents regions where the model is confident (beyond the user-provided confidence threshold) that carbon storage will either increase or decrease in the transition from the current scenario to the future baseline scenario. The second raster represents regions where the model is confident that carbon storage will either increase or decrease in the transition from the current scenario to the REDD policy scenario.

The model currently does not support REDD scenario analysis together with harvested wood product analysis. Therefore, if REDD scenario analysis is enabled, HWP analysis will be disabled.

Limitations and Simplifications

The model greatly oversimplifies the carbon cycle which allows it to run with relatively little information, but also leads to important limitations. For example, the model assumes that none of the LULC types in the landscape are gaining or losing carbon over time. Instead it is assumed that all LULC types are at some fixed storage level equal to the average of measured storage levels within that LULC type. Under this assumption, the only changes in carbon storage over time are due to changes from one LULC type to another or from the harvest of wood products. Therefore, any grid cell that does not change its LULC type and is at a wood harvest steady-state will have a sequestration value of 0 over time. In reality, many areas are recovering from past land use or are undergoing natural succession. The problem can be addressed by dividing LULC types into age classes (essentially adding more LULC types), such as three ages of forest. Then, parcels can move from one age class to the other in scenarios and change their carbon storage values as a result.

A second limitation is that because the model relies on carbon storage estimates for each LULC type, the results are only as detailed and reliable as the LULC classification used. Carbon storage clearly differs among LULC types (e.g., tropical forest vs. open woodland), but often there can also be significant variation within a LULC type. For example, carbon storage within a "tropical moist forest" is affected by temperature, elevation, rainfall, and the number of years since a major disturbance (e.g., clear-cut or forest fire). The variety of carbon storage values within coarsely defined LULC types can be partly recovered by using a LULC classification system and related carbon pool table which stratifies coarsely defined LULC types with relevant environmental and management variables. For example, forest LULC types can be stratified by elevation, climate bands or time intervals since a major disturbance. Of course, this more detailed approach requires data describing the amount of carbon stored in each of the carbon pools for each of the finer LULC classes.

Another limitation of the model is that it does not capture carbon that moves from one pool to another. For example, if trees in a forest die due to disease, much of the carbon stored in aboveground biomass becomes carbon stored in other (dead) organic material. Also, when trees are harvested from a forest, branches, stems, bark, etc. are left as slash on the ground. The model assumes that the carbon in wood slash "instantly" enters the atmosphere.

With respect to its estimates of carbon in HWPs, the model is constrained by the fact that users may assign only one harvest rate (e.g., 50 Mg of wood per harvest where a harvest occurs every 2 years) and only one decay rate (e.g., the wood harvested from the parcel over the years is always used to make the same product that decays at the same rate) to each parcel. In reality, harvested parcels will exhibit variation in harvest and decay rates over time. The model also does not account for the greenhouse gasses (GHGs) emitted from the transportation of harvested wood from its initial harvest site to its final destination, the conversion of raw wood into finished products, or agriculture-related activities such as from tractors and livestock. Annual GHG emissions from agricultural land use can be calculated with the InVEST Agriculture Production Model, due to be released soon.

The uncertainty model has a few limitations. First, it assumes that the probability distribution for amount of carbon in different pools is normally distributed. This may not be the case; for instance, predictions for carbon amounts may be asymmetric distributions. If this is the case, users should choose a normal distribution that best approximates the desired distribution. In addition, uncertainty in input data is currently limited to carbon pools. The model does not yet handle uncertainty in LULC maps or HWP data. The carbon model also assumes that every carbon pool across different LULC types is independent. If, in reality, estimates for carbon pools are consistently too high or too low, then error may be greater than predicted by the model.

Finally, while most sequestration follows a nonlinear path such that carbon is sequestered at a higher rate in the first few years and a lower rate in subsequent years, the model's economic valuation of carbon sequestration assumes a linear change in carbon storage over time. The assumption of a constant rate of change will tend to undervalue the carbon sequestered, as a nonlinear path of carbon sequestration is more socially valuable due to discounting than a linear path (Fig.2).

Figure 2: The model assumes a linear change in carbon storage (the solid line), while the actual path to the year T's carbon storage level may be non-linear (like the dotted line). In this case t can indicate the year of the current landscape and T the year of the future landscape. With positive discounting, the value of the modeled path (the solid line) is less valuable than the actual path. Therefore, if sequestration paths tend to follow the dotted line, the modeled valuation of carbon sequestration will underestimate the actual value of the carbon sequestered.

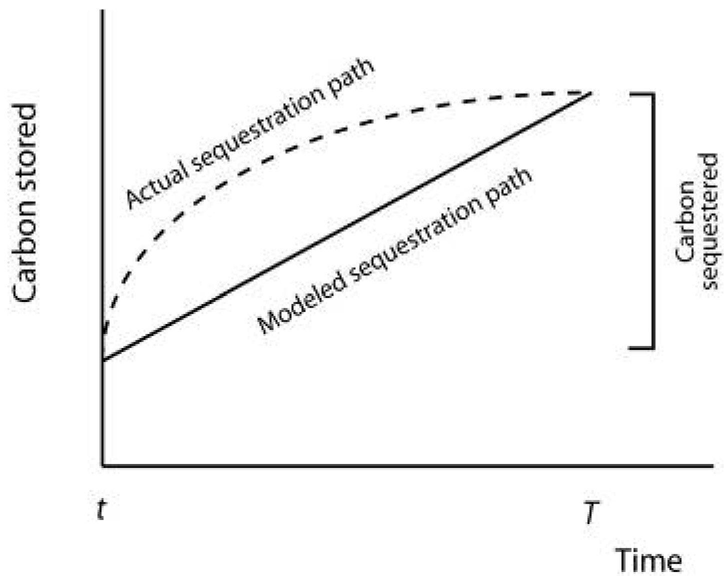
5.2.4 Data Needs

This section outlines the map and data tables required by the model, including the economic data that the tool interface will prompt the user to enter. See Appendix for detailed information on data sources and pre-processing.

1. **Current land use/land cover (LULC) map (required):** A GIS raster dataset, with a LULC code for each cell. The dataset should be projected in meters and the projection used should be defined.

   Name: file can be named anything, but avoid spaces

   Format: standard GIS raster file (e.g., ESRI GRID or IMG), with LULC class code for each cell (e.g., 1 for forest, 3 for grassland, etc.) These codes must match LULC codes in the tables below. LULC class



codes should be in the 'LULC' column of the dataset.

Sample data set: \Invest\Base_Data\Terrestrial\lulc_samp_cur

The model requires the following two pieces of information about the LULC map which are prompted for in the interface.

- The **year** depicted by the LULC map, for use in calculating sequestration and economic values (labeled "Year of current land cover" in the interface).
- The **spatial resolution** (desired cell size in meters) at which you would like the model to run (labeled "Resolution (optional)"). You can only define a new resolution that is coarser than the resolution of the LULC map (this is the default resolution).

2. **Carbon pools (required):** A table of LULC classes, containing data on carbon stored in each of the four fundamental pools for each LULC class. Carbon storage data can be collected from field estimates from local plot studies, extracted from meta-analyses on specific habitat types or regions, or found in general published tables (e.g., IPCC, see Appendix). If information on some carbon pools is not available, pools can be estimated from other pools, or omitted by leaving all values for the pool equal to 0.

   If a forest is regularly harvested for woody biomass, the estimates of carbon biomass in the aboveground, belowground, and dead organic matter pools should reflect this fact. For example, suppose one of the LULC types is a plantation forest that tends to have one-tenth of its area clear-cut every year. The aboveground and belowground estimates of carbon biomass for this LULC type should reflect the fact that only 9/10ths of the area occupied by plantation forests will be covered by trees at any point in time.

   For notes on calculating standard deviation for the uncertainty model, see the Appendix for data sources for carbon stocks.

   Name: file can be named anything

   File type: *.csv or *.dbf

   Rows: each row is a LULC class

   Columns: each column contains a different attribute of each LULC class, and must be named as follows:

   - lucode: code of land use/land cover class (e.g., 1 for forest, 3 for grassland, etc.). The LULC code should match the LULC codes from the current LULC map (dataset #1 above)
   - LULC_name: descriptive name of LULC class (optional)

To run the model **without uncertainty analysis**, the following columns are required:

- C_above: amount of carbon stored in aboveground biomass (in Mg ha^{-1})
- C_below: amount of carbon stored in belowground biomass (in Mg ha^{-1})
- C_soil: amount of carbon stored in soil (in Mg ha^{-1})
- C_dead: amount of carbon stored in dead organic matter (in Mg ha^{-1})

To run the model **with uncertainty analysis**, the following columns are required:

- C_above_mean: estimated amount of carbon stored in aboveground biomass (in Mg ha^{-1})
- C_above_sd: standard deviation to measure uncertainty in the amount of carbon in aboveground biomass (in Mg ha^{-1})
- C_below_mean: estimated amount of carbon stored in belowground biomass (in Mg ha^{-1})
- C_below_sd: standard deviation to measure uncertainty in the amount of carbon in belowground biomass (in Mg ha^{-1})
- C_soil_mean: estimated amount of carbon stored in soil (in Mg ha^{-1})
- C_soil_sd: standard deviation to measure uncertainty in the amount of carbon in soil (in Mg ha^{-1})
- C_dead_mean: estimated amount of carbon stored in dead organic matter (in Mg ha^{-1})
- C_dead_sd: standard deviation to measure uncertainty in the amount of carbon in dead organic matter (in Mg ha^{-1})

Note: The unit for all carbon pools is Mg of elemental carbon ha^{-1}. This means that if your data source has information on Mg of CO_2 stored ha^{-1}, you need to convert those numbers to elemental carbon by multiplying Mg of CO_2 stored $ha^{-1}$by 0.2727.

Sample data set (without uncertainty): \Invest\Carbon\Input\carbon_pools_samp.csv

Sample data set (with uncertainty): \Invest\Carbon\Input\carbon_pools_samp_uncertain.csv

Example (without uncertainty): Hypothetical study with five LULC classes. Class 1 (Forest) contains the most carbon in all pools. In this example, carbon stored in above- and below-ground biomass differs strongly among land use classes, but carbon stored in soil varies less dramatically.

lucode	LULC_name	C_above	C_below	C_soil	C_dead
1	Forest	140	70	35	12
2	Coffee	65	40	25	6
3	Pasture/grass	15	35	30	4
4	Shrub/undergrowth	30	30	30	13
5	Open/urban	5	5	15	2

Example (with uncertainty): As above, but with standard deviations to measure uncertainty in carbon pool estimates.

lucode	LULC_name	C_above_mean	C_above_sd	C_below_mean	C_below_sd	C_soil_mean	C_soil_sd	C_dead_mean	C_dead_sd
1	Forest	140	20	70	10	35	5	12	2
2	Coffee	65	5	40	10	25	5	6	2
3	Pasture/grass	15	3	35	5	30	5	4	1
4	Shrub/undergrowth	30	5	30	7	30	8	13	3
5	Open/urban	5	1	5	1	15	2	2	0.5

Confidence threshold (for uncertainty): The uncertainty model also requires an additional *confidence threshold* parameter which is input directly through the tool interface rather than through a file. This is used as the minimum probability of storage increase or decrease for which we highlight cells in the 'conf' output file.

3. **Current harvest rates map (optional)**. A GIS shape file of polygons (parcels in our vernacular), contains data on:

   1. Parcel ID
   2. Amount of carbon, in the form of woody biomass, typically removed from the parcel over the course of a harvest period
   3. Date that the modeler wants to begin accounting for wood harvests in the parcel
   4. Frequency of harvest periods in the parcel in the past
   5. Average decay rate of products made from the wood harvested from a parcel
   6. Average carbon density of the wood removed form the parcel in the past
   7. Average tree volume per ton of wood removed form the parcel in the past.

   The GIS polygon map should only delineate parcels that have been harvested; all other portions of the landscape should be ignored. Note that unlike the current LULC map,this file contains multiple data for each individual harvest parcel on the landscape.

   The amount of carbon that is removed, on average, during each harvest period can be estimated from plot surveys, market demand analyses, community surveys, or based on expert opinion. Decay rates can be estimated from literature reports (see sources in Appendix) or also based on expert opinion if necessary. If multiple types of wood products are harvested from a polygon, the user should average the rates of decay or focus on the product with the slowest decay rate (since that will affect storage the most). Because only woody biomass is included in the harvest portion of the model, it is not necessary to include harvest or decay rates for herbaceous products. If you are unable or uninterested in estimating carbon stored in harvested wood products, you do not need to supply this table and the model will ignore this pool.

   Name: file can be named anything

   File type: GIS polygon shapefile

   Rows: each row is a specific polygon on the landscape.

   Columns: columns contain attributes related to harvested wood products and must be named as follows:

   1. FID: unique identifying code for each polygon (parcels in our vernacular).
   2. *Cut_cur*: The amount of carbon typically removed from a parcel during a harvest period (measured in Mg ha^{-1}; the model will sum across the area of each parcel). This amount should only include the portion of the wood's carbon that is removed from the parcel (e.g., the carbon in the wood delivered to a saw mill). In other words, the slash and other waste from a wood harvest should be ignored because the model assumes that its carbon content is lost to the atmosphere instantly (the "cur" at the end of this attribute is used to relate it to the "current" LULC map).
   3. *Start_date*: The first year the carbon removed from a forest will be accounted for in the HWP pool. The first year should coincide with a year in which wood was actually harvested from the parcel. If wood was harvested from a parcel in 1995, 2000, and 2005 and the LULC map being evaluated is from 2005 then St_date can equal 1995, 2000, or 2005; it is your choice.
   4. *Freq_cur*: The frequency, in years, with which the Cut_cur amount is harvested. If the value is 1 then the Cut_cur amount is removed annually from the parcel, if 5 then every 5 years, etc.
   5. *Decay_cur*: The half-life of wood products harvested, measured in years.
   6. *C_den_cur*: The carbon density in the harvested wood (MgC $Mg^{-1}$of dry wood). Typically, the statistic ranges between 0.43 and 0.55 (see table 4.3 of IPCC (2006)). If C_den_cur is not known for a parcel set it equal to 0.5.

7. *BCEF_cur*: An expansion factor that translates the mass of harvested wood into volume of harvested wood (Biomass Conversion Expansion Factor). The expansion factor is measured in Mgof dry wood per m3 of wood and is a function of stand type and stand age. If you do not have data on this expansion factor you can use the BCEFR row in table 4.5 of IPCC (2006). Otherwise, set this expansion factor equal to 1 for each parcel.

Sample data set: \Invest\Carbon\Input\harv_samp_cur.shp

Example: A hypothetical study of carbon storage in HWP for four forest parcels that have experienced harvests in the past. Assume the current LULC map we are using corresponds to the year 2005. Parcels 1, 2, and 3 are forests that are managed for timber production. Each managed forest experiences a cut every 5th year where Cut_cur gives the amount of carbon (Mg ha^{-1}) in the portion of the wood that is removed every fifth year. The fourth parcel is a source of firewood and wood is cut from the parcel continuously. Thus, for this parcel we estimate the annual rate of carbon removed from the forest for firewood. For the first three parcels, we began to account for carbon removal in 1995. For the final parcel we began accounting for HWP in 2000. (Recall that the calculation of HWP_cur, Bio_HWP_cur, and Vol_HWP_cur does not include the 2005 harvest; that carbon is still on the land.)

FID	Cut_cur	Start_date	Freq_cur	Decay_cur	C_den_cur	BCEF_cur
1	75	1995	5	30	0.5	1
2	50	1995	5	35	0.5	1
3	50	1995	5	50	0.5	1
4	45	2000	1	1	0.5	1

We measure the carbon stored in HWP that originated from parcel x on the current landscape with the following equation:

$$HWP\_cur_x = Cut\_cur_x \times \sum_{t=0}^{ru\left(\frac{yr\_cur-start\_date}{Freq\_cur_x}\right)-1} f(Decay\_cur_x; yr\_cur - start\_date_x - (t \times Freq\_cur_x)) \tag{5.1}$$

where HWP_curx is measured in Mg ha^{-1}, yr_cur is short for "Year of current land cover", t indexes the number of harvest periods, and ru indicates that any fraction should be rounded up to the next integer value. The function

$$f(\bullet) = \left\lfloor \frac{1-e^{-\omega_x}}{\omega_x \times e^{[yr\_cur-start\_date_x-(t\times Freq\_cur_x)]\times\omega_x}} \right\rfloor \tag{5.2}$$

where $\omega_x = (\log_e 2/Decay\_cur_x)$, measures how much of the carbon was typically removed from a parcel (Cut_curx) during a harvest period, that occurred some number of years ago ($yr\_cur - start\_date_x - (t \times Freq\_cur_x)$), still remains trapped in HWP as of the current year (yr_cur) and given the current decay rate ($Decay\_curx$).

The following are several examples to show how equation (1) works. In the first instance, assume $start\_datex = 1983$, $yr\_cur = 2000$, and $Freq\_curx = 4$. In this case, $ru\left(\frac{yr\_cur-start\_date}{Freq\_cur_x}\right) = ru\left(\frac{17}{4}\right) = ru(4.25) = 5$. According to the summation term in equation (1), this means we sum over 5 harvest periods (t = 0,1,2,3,4). Given this series of t, we evaluate f at 17, 13, 9, 5, and 1 years since a harvest (we use to convert the series of t's into years since harvest).

Alternatively, if $start\_datex = 1980$, $yr\_cur = 2000$, and $Freq\_curx = 2$ then $ru\left(\frac{yr\_cur-start\_date}{Freq\_cur_x}\right) = ru(10) = 10$. Therefore, according to equation (1), harvests that contained Cut_curx of carbon ha^{-1} occurred on the parcel 20, 18, 16, 14, 12, 10, 8, 6, 4, and 2 years before the year 2000 (note that we do not include a harvest that is scheduled to occur in the current year in the HWP carbon pool; this carbon is still in situ in the current year).

We use C_den_cur and $BCEF\_cur$ to measure the mass (Bio_HWP_cur) and volume (Vol_HWP_cur) of wood that has been removed from a parcel from the $start\_date$ to the current year. Bio_HWP_curfor parcel x is measured in Mg (dry matter) ha^{-1} and is given by:

$$Bio\_HWP\_cur_x = Cut\_cur_x \times ru\left(\frac{yr\_cur - start\_date}{Freq\_cur_x}\right) \times \frac{1}{C\_den\_cur_x} \tag{5.3}$$

and Vol_HWP_cur for parcel x is measured in m^3 of wood ha^{-1} and is given by,

$$Vol\_HWP\_cur_x = Bio\_HWP\_cur_x \times \frac{1}{Vol\_exp\_cur_x} \tag{5.4}$$

As mentioned before, the model places all parcel-level values into a grid cell map that comports with the four pool storage map.

4. **Future Scenarios (optional – required for valuation)**: If you have a LULC map (data input #1) for a future landscape scenario, then expected sequestration rates in the four major carbon pools on the landscape can be measured. Similarly, sequestration rates in the HWP carbon pool can be measured with a harvest rate map (data input #3) for this future landscape.

   If REDD scenario analysis is enabled, then this should represent the landscape for the future baseline scenario, against which the REDD scenario will be compared.

   A future land cover map (a raster dataset) should be formatted according to the same specifications as the current land cover map (input #1).

   If you provide a future harvest rate map then the HWP carbon pool can be tracked over time. The future harvest rate map should be formatted according to the same specifications as the current harvest rate map: a polygon map where values for *FID*, *Cut_fut*, *Freq_fut*, *Decay_fut*, *C_den_fut*, and *BCEF_fut* are attributed to each parcel that is expected be harvested at some point between the year given by $\frac{yr\_cur+yr\_fut}{2}$ and *yr_fut* where *yr_fut* indicates the year associated with the future land cover map (e.g., if *yr_cur* is 2000 and *fut_yr* is 2050 then $\frac{yr\_cur+yr\_fut}{2}$ = 2025). This means that current harvest rate map conditions hold on the landscape until the year halfway between the current and future years. The harvest variables for the future will be applied in the year $\frac{yr\_cur+yr\_fut}{2}$. Note that any fraction is rounded up (e.g., if *yr_cur* is 2000 and *fut_yr* is 2053 then $\frac{yr\_cur+yr\_fut}{2}$ = 2026). The future harvest rate map does not have to retain any spatial semblance to the current harvest rate map. Nor do parcels that are harvested on the current and future maps have to have a common FID.

   Sample data files for future scenarios are future land cover: (\InVEST\Base_Data\Terrestrial\lulc_samp_fut) and future harvest rate map (\InVEST\Carbon\Input\harv_samp_fut.shp).

   Example: A hypothetical study of future carbon storage in HWP for four forest parcels. Continuing with current harvest rate map (2005) described above, assume the future LULC map corresponds to the year 2035. Three of the four forest parcels that have wood removed on the current landscape keep their boundaries in the future and continue to have wood removed into the future (parcels with FID 1, 3, and 4 on the current harvest rate map). However the first parcel changes its management with *newCut* and *Freq* values ($Cut\_cur_x \neq Cut\_fut_x$ and $Freq\_cur_x \neq Freq\_fut_x$). We assume these new management conditions begin in the year 2020 (given by $\frac{yr\_cur+yr\_fut}{2}$). Parcel 2 is not expected to be harvested at any point between $\frac{yr\_cur+yr\_fut}{2}$ and *yr_fut*. Therefore, the model assumes that the harvest activity given in current harvest rate map for parcel 2 ends in 2020. In addition, the future harvest rate map includes a new harvested parcel (given by FID = 5). We assume that harvest begins there in 2020 as well. In parcels 3 and 4 harvest management does not change across the current and future landscapes. (Note that we retained the FID values across the two maps here; this is not necessary, as the ArcGIS program will perform the necessary spatial matches).

FID	Cut_fut	Freq_fut	Decay_fut	C_den_fut	BCEF_fut
1	50	10	30	0.5	1
3	50	5	50	0.5	1
4	45	1	1	0.5	1
5	25	2	15	0.5	1

Below we describe exactly how the future harvest values are calculated. If a parcel was harvested on the current landscape and is expected to be harvested on the future landscape (i.e., at some point between $\frac{yr\_cur+yr\_fut}{2}$ and yr_fut) then the remaining HWP carbon due to harvest from parcel x in the future year is given by:

$$\begin{aligned}HWP\_fut_x = \quad & Cut\_cur_x \sum_{t=0}^{ru\left(\frac{\frac{yr\_fut+yr\_cur}{2}-start\_date_x}{Freq\_cur_x}\right)-1} f(Decay\_cur_x, yr\_fut - start\_date_x - (t \times Freq\_cur_x)) + \\ & Cut\_fut_x \sum_{t=0}^{ru\left(\frac{yr\_fut-\frac{yr\_fut+yr\_cur}{2}}{Freq\_fut_x}\right)-1} f\left(Decay\_fut_x, yr\_fut - \frac{yr\_fut+yr\_cur}{2} - (t \times Freq\_fut_x)\right)\end{aligned} \tag{5.5}$$

where the function f is as before. Recall that if (yr_cur + yr_fut) / 2 results in a fraction it is rounded up. Also note that equation (5) does not include a harvest that is scheduled to occur in the future year; this harvest's carbon isin situ in this accounting. Parcels that were harvested on the current landscape but are not expected to be harvested on the future landscape may still have HWP carbon in the future year. The remaining HWP carbon in yr_fut on such parcels is given by the first term of equation (5):

$$HWP\_fut_x = Cut\_cur_x \times \sum_{t=0}^{ru\left(\frac{\frac{yr\_fut+yr\_cur}{2}-start\_date_x}{Freq\_cur_x}\right)-1} f(Decay\_cur_x, yr\_fut - start\_date_x - (t \times Freq\_cur_x)) \tag{5.6}$$

In contrast, parcels that were not harvested on the current landscape, but are expected to be harvested on the future landscape, will have the following amount of carbon in the form of HWP in yr_fut:

$$HWP\_fut_x = Cut\_fut_x \sum_{t=0}^{ru\left(\frac{yr\_fut-\frac{yr\_fut+yr\_cur}{2}}{Freq\_fut_x}\right)-1} f\left(Decay\_fut_x, yr\_fut - \frac{yr\_fut+yr\_cur}{2} - (t \times Freq\_fut_x)\right) \tag{5.7}$$

Note that this is the second term of equation (5).

If a parcel was harvested on the current landscape and is expected to be harvested on the future landscape, the mass of harvested wood that has been removed from a parcel from Start_date to yr_fut is given by:

$$\begin{aligned}Bio\_HWP\_fut_x = \quad & \left(Cut\_cur_x \times ru\left(\frac{\frac{yr\_fut+yr_cur}{2}-start\_date_x}{Freq\_cur_x}\right) \times \frac{1}{C\_den\_cur_x}\right) + \\ & \left(Cut\_fut_x \times ru\left(\frac{yr\_fut-\frac{yr\_fut+yr\_cur}{2}}{Freq\_fut_x}\right) \times \frac{1}{C\_den\_fut}\right)\end{aligned} \tag{5.8}$$

However, for parcels that were harvested on the current landscape, but are not expected to be harvested on the future landscape, the mass of wood removed from a parcel from *Start_date* to *yr_fut* is given by the first term of equation (8):

$$Bio\_HWP\_fut_x = \left(Cut\_cur_x \times ru\left(\frac{\frac{yr\_fut+yr_cur}{2} - start\_date_x}{Freq\_cur_x}\right) \times \frac{1}{C\_den\_cur_x}\right) \tag{5.9}$$

For parcels that were not harvested on the current landscape but are expected to be harvested on the future landscape, the mass of wood removed from a parcel from Start_date toyr_futis given by second term of equation (8):

$$Bio\_HWP\_fut_x = \left(Cut\_fut_x \times ru\left(\frac{yr\_fut - \frac{yr\_fut+yr\_cur}{2}}{Freq\_fut_x}\right) \times \frac{1}{C\_den\_fut}\right) \tag{5.10}$$

Finally, the volume of the of wood that has been removed from a parcel from *Start_date* to *yr_fut* is given by:

$$\begin{aligned} Vol\_HWP\_fut_x = \ & \left(Cut\_cur_x \times ru\left(\frac{\frac{yr\_fut+yr\_cur}{2} - start\_date_x}{Freq\_cur_x}\right) \times \frac{1}{C\_den\_cur_x} \times \frac{1}{BCEF\_cur_x}\right) + \\ & \left(Cut\_fut_x \times ru\left(\frac{yr\_fut - \frac{yr\_fut+yr\_cur}{2}}{Freq\_fut_x}\right) \times \frac{1}{C\_den\_fut_x} \times \frac{1}{BCEF\_fut_x}\right) \end{aligned} \tag{5.11}$$

$$Vol\_HWP\_fut_x = \left(Cut\_cur_x \times ru\left(\frac{\frac{yr\_fut+yr\_cur}{2} - start\_date_x}{Freq\_cur_x}\right) \times \frac{1}{C\_den\_cur_x} \times \frac{1}{BCEF\_cur_x}\right) \tag{5.12}$$

or

$$Vol\_HWP\_fut_x = \left(Cut\_fut_x \times ru\left(\frac{yr\_fut - \frac{yr\_fut+yr\_cur}{2}}{Freq\_fut_x}\right) \times \frac{1}{C\_den\_fut_x} \times \frac{1}{BCEF\_fut_x}\right) \tag{5.13}$$

depending on the combination of current and future harvests (see above).

We recommend that the modeler use *Bio_HWP_cur* and *Bio_HWP_fut* to refine the current and future LULC maps. Specifically, if *Bio_HWP_cur* or *Bio_HWP_fut* on a portion of the landscape are significant, then the modeler should assess whether the LULC types associated with that portion of the current or future landscape accurately reflect the biomass remaining on the landscape. For example, if the current LULC type on a portion of the landscape that has been heavily harvested in the immediate past is “closed conifer” it may be more appropriate to reclassify it as “thinned conifer” or “open conifer” on the LULC map.

5. **REDD scenario LULC map (optional)**. REDD scenario analysis requires a LULC map for a landscape scenario under a REDD policy. This should be formatted according to the same specifications as the current and the baseline future land cover map. The REDD scenario LULC map must be for the same year as the baseline future scenario LULC map.

6. **Economic data (optional – required for valuation)**. Three numbers are not supplied in a table, but instead are input directly through the tool interface.

   (a) The **value of a sequestered ton of carbon** (*V* in the equation below), in dollars per metric ton of elemental carbon (not CO_2, which is heavier, so be careful to get units right! If the social value of CO_2e is \$Y per metric ton, then the social value of C is \$(3.67*Y) per metric ton (Labeled “Price of carbon per metric ton (optional)” in the tool interface.) For applications interested in estimating the total value of carbon sequestration, we recommend value estimates based of damage costs associated with the release of an additional ton of carbon (the social cost of carbon (SCC). Stern (2007), Tol (2009), and Nordhaus (2007a) present estimates of SCC. For example, two SCC estimates we have used from Tol (2009) are \$66 and \$130 (in 2010 US dollars) (Polasky et al. 2010). For applications interested in estimating the value that could be gained by trading carbon credits in the current markets, the value can be taken from the current market prices on the Chicago or European Climate Exchanges.

(b) The **market discount rate** (*r* in the equation below), which reflects society's preference for immediate benefits over future benefits (labeled "Market discount rate (%) (optional)" in the tool interface). The default value in the interface is 7% per year, which is one of the market discount rates recommended by the U.S. government for cost-benefit evaluation of environmental projects. However, this rate will depend on the country and landscape being evaluated. Philosophical arguments have been made for using a lower discount rate when modeling climate change related dynamics, which users may consider using. If the rate is set equal to 0% then monetary values are not discounted.

(c) The **annual rate of change in the price of carbon** (*c* in the equation below), which adjusts the value of sequestered carbon as the impact of emissions on expected climate change-related damages changes over time. The default value in the interface is 0% (labeled "The annual rate of change in the price of carbon (%) (optional)" in the tool interface). However, settingthis rate greater than 0% suggests that the societal value of carbon sequestered in the future is less than the value of carbon sequestered now. It has been widely argued that GHG emissions need to be curtailed immediately to avoid crossing a GHG atmospheric concentration threshold that would lead to a 3 degree Celsius or greater change in global average temperature by 2105.Some argue that such a temperature change would lead to major disruptions in economies across the world (Stern et al. 2006). Therefore, any mitigation in GHG emissions that occurs many years from now may have no effect on whether or not this crucial concentration threshold is passed. If this is the case, C sequestration in the far future would be relatively worthless and a carbon discount rate greater than zero is warranted. Alternatively, setting the annual rate of change less than 0% (e.g., -2%) suggests that the societal value of carbon sequestered in the future is greater than the value of carbon sequestered now (this is a separate issue than the value of money in the future, a dynamic accounted for with the market discount rate). This may be the case if the damages associated with climate change in the future accelerate as the concentration of GHGs in the atmosphere increases.

The value of carbon sequestration over time for a given parcel *x* is:

$$value\_seq_x = V\frac{sequest_x}{yr\_fut - yr\_cur}\sum_{t=0}^{yr\_fut-yr\_cur-1}\frac{1}{\left(1+\frac{r}{100}\right)^t\left(1+\frac{c}{100}\right)^t} \tag{5.14}$$

5.2.5 Running the Model

The model is available as a standalone application accessible from the Windows start menu. For Windows 7 or earlier, this can be found under *All Programs -> InVEST +VERSION+ -> Carbon*. Windows 8 users can find the application by pressing the windows start key and typing "carbon" to refine the list of applications. The standalone can also be found directly in the InVEST install directory under the subdirectory *invest-3_x86/invest_carbon.exe*.

Viewing Output from the Model

Upon successful completion of the model, a file explorer window will open to the output workspace specified in the model run. This directory contains an *output* folder holding files generated by this model. Those files can be viewed in any GIS tool such as ArcGIS, or QGIS. These files are described below in Section *Interpreting Results*.

5.2.6 Interpreting Results

Final Results

Final results are found in the *Output* folder within the *Workspace* specified for this module.

If REDD scenario analysis is enabled, then files with the suffix *_base* represent results for the baseline future scenario, and files with the suffix *_redd* represent results for the REDD policy scenario.

Model results:

- **summary.html:** This file presents a summary of all data computed by the model. It also includes descriptions of all other output files produced by the model, so it is a good place to begin exploring and understanding model results. Because this is an HTML file, it can be opened with any web browser.
- **Parameter log**: Each time the model is run, a text (.txt) file will appear in the *Output* folder. The file will list the parameter values for that run and will be named according to the service, the date and time, and the suffix.
- **tot_C_cur:** This file shows the amount of carbon currently stored in Mg in each grid cell at the chosen resolution. This is a sum of all of the carbon pools you have included data for (above ground, below ground, soil, dead material, and harvested wood product). The lowest value can be 0 (for example, paved areas if you don't include the soil beneath the pavement). Examine this map to see where high and low values fall. Is this what you would expect given the current land use and land cover? If not, check your input files.
- **tot_C_fut:** This file shows the total amount of carbon that will be stored in each parcel under your future landscape scenario. It is a sum of all the carbon pools for which you have included data. The values are in Mg per grid cell. Again, the lowest value can be 0.
- **sequest:** This file maps the difference in carbon stored between the future landscape and the current landscape – or the carbon that is sequestered during the entire given time period (i.e. this is a rate per the total time period elapsed, yr_fut – yr_cur, not per year). The values are in Mg per grid cell. In this map some values may be negative and some positive. Positive values indicate sequestered carbon, whereas negative values indicate carbon that was lost. Areas with large negative or positive values should have the biggest changes in LULC or harvest rates. Remember that carbon emissions due to management activities (tractors burning fuel, fertilizer additions, etc.) on a parcel are NOT included in this assessment.
- **conf** *(for uncertainty model only)*: This file maps areas where we are confident that emissions either increase or decrease. Grid cells where we are confident that storage will increase from the current LULC map to the future LULC map have a value of 1. Grid cells where we are confident storage will decrease have a value of -1. Grid cells where we are not confident either way have a value of 0. The confidence threshold specified by the user in the initial parameters is used as the minimum probability threshold for which we highlight a region with a 1 or -1. For example, if the user specifies a confidence threshold of 95, a grid cell will receive a value of 1 only if it is at least 95% likely that storage will increase in that particular cell.
- **value_seq:** This file maps the economic value of carbon sequestered (between the current and the future landscape dates, yr_cur and yr_fut). The relative differences between parcels should be similar (but not identical) to sequest, but the values are in dollars per grid cell instead of Mg per grid cell. As with sequest, values may be negative, indicating the cost of carbon emissions from LULC changes to that parcel.
- ***_mask files** *(for uncertainty model only)*: When provided with confidence rasters, the valuation model will produce files such as **seq_mask** and **val_mask**. These files contain the raster created by 'masking' the **sequest** and **value_seq** rasters, respectively, with the **conf** confidence raster. In other words, **seq_mask** is identical to **sequest**, except that areas where the **conf** raster indicates low confidence are ignored (and set to 'no data' values). Similarly, **val_mask** is identical to **value_seq**, except that areas where the **conf** raster indicates low confidence are ignored. Therefore, the ***mask** files contain values only in those cells where we have high confidence that carbon storage will increase or decrease.

Intermediate Results

These files independently map each of the five carbon pools that contribute to the final results for both current and future landscapes. Examining these results can help you determine which of the carbon pools are changing the most between your current and future landscapes and can help you identify areas where your data may need correcting. The unit for each of these pool outputs is Mg per grid cell. *Biomass_HWP_cur* and *Biomass_HWP_fut* are both measured in Mg dry matter per grid cell and *Vol_HWP_cur* and *Vol_HWP_fut* are both measured in m^3 of wood per grid cell.

lc_res_cur and *lc_res_fut* give the current and future LULC maps at the resolution chosen with the model interface. Finally, Carbon_dateandtime_suffix.txt is a text file that summarizes the parameter data you chose when running the Carbon Storage and Sequestration Model. The text file's name includes "dateandtime" which means that the data and time is stamped into the text's file name. The text file's name also includes a "suffix" term that you choose.

- *C_above_cur* – the current carbon stock for the aboveground pool
- *C_above_fut* – the carbon stock for the aboveground pool for the future scenario
- *C_below_cur* – the current carbon stock for the belowground pool
- *C_below_fut* – the carbon stock for the belowground pool for the future scenario
- *C_soil_cur* – the current carbon stock in soil
- *C_soil_fut* – the carbon stock in soil for the future scenario
- *C_dead_cur* – the current carbon stock in dead organic matter
- *C_dead_fut* – the carbon stock in dead organic matter for the future scenario
- *C_HWP_cur* – carbon stored in harvested wood products for current land cover
- *C_HWP_fut* – carbon stored in harvested wood products for future scenario.
- *Bio_HWP_cur* – biomass of wood removed since "start_date" for current land cover
- *Bio_HWP_fut* – biomass of wood removed since "start_date" for future land cover
- *Vol_HWP_cur* – volume of wood removed since "start_date" for current land cover
- *Vol_HWP_fut* – volume of wood removed since "start_date" for future land cover
- *lc_res_cur* – the current LULC map at the resolution chosen by the user.
- *lc_res_fut* – the future LULC map at the resolution chosen by the user.
- *Carbon_dateandtime_suffix.txt* – a text file that summarizes the parameter data used to run the Carbon Storage and Sequestration Model.

5.2.7 Appendix: Data Sources

This is a rough compilation of data sources and suggestions for finding, compiling, and formatting data. This section should be used for ideas and suggestions only. This section is updated as new data sources and methods become available.

1. Land use/land cover map

The simplest categorization of LULCs on the landscape involves delineation by land cover only (e.g., cropland, temperate conifer forest, prairie). Several global and regional land cover classifications are available (e.g., Anderson et al. 1976), and often detailed land cover classification has been done for the landscape of interest.

A slightly more sophisticated LULC classification could involve breaking relevant LULC types into broad age categories (e.g., forest of age 0-10 years, 11-20, 21-40, etc.). This would allow separate estimates of carbon storage for different ages. In scenarios, parcels can move from one age class to the next, crudely capturing changes in carbon storage over time. This approach requires more information, however, including carbon storage estimates for each age class for all modeled pools of carbon.

A still more detailed classification could stratify LULC types by variables known to affect carbon storage within a given LULC type (e.g., montane forest 800-1000m, montane forest 1001-1200m, etc.). Rainfall, temperature, and elevation all typically influence carbon storage and sequestration (e.g., Jenny 1980, Coomes et al. 2002, Raich et al. 2006). If data are available to estimate carbon storage at different elevations, or at different levels of rainfall, temperature or other climate variables, model results will be substantially more accurate. This will typically take a large sample of plot estimates of carbon storage.

2. Carbon stocks

Carbon storage data should be set equal to the average carbon storage values for each LULC class. The ideal data source for all carbon stocks is a set of local field estimates, where carbon storage for all relevant stocks has been directly measured. These can be summarized to the LULC map, including any stratification by age or other variable. If these data are not available, however, there are several general data sources that can be used.

Note that several sources, including IPCC (2006), report in units of biomass, while InVEST uses mass of elemental carbon. To convert metric tons of biomass to metric tons of C, multiply by a conversion factor, which varies typically from 0.43 to 0.51. Conversion factors for different major tree types and climatic regions are listed in Table 4.3 on page 4.48 of IPCC (2006).

Notes on calculating standard deviation for the uncertainty model: The standard deviation values in the carbon pool table signify uncertainties in the true value for amount of carbon in different pools. There are a variety of methods to calculate standard deviation. For instance, the standard deviation can be calculated from a confidence interval; a 95 percent confidence interval, for example, is 3.92 standard deviations wide. Therefore, we can divide the width of the 95 percent confidence interval by 3.92 to calculate standard deviation. For more information on uncertainty analysis, see Volume 1 Chapter 3, "Uncertainties", in IPCC (2006).

2.1. Carbon stored in aboveground biomass

A good but very general source of data for carbon storage is the Intergovernmental Panel on Climate Change's (IPCC) 2006 methodology for determining greenhouse gas inventories in the Agriculture, Forestry and Other Land Use (AFOLU) sector (http://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/4_Volume4/V4_02_Ch2_Generic.pdf, IPCC 2006). To use this set of information from the IPCC, you must know your site's climate domain and region; use data from Table 4.1 on page 4.46 and a digital copy of the Food and Agriculture Organization of the United Nations' (FAO) eco-region map (http://www.fao.org/geonetwork/srv/en/main.home) to figure that out. Tables 5.1 through 5.3 (p. 5.9) of IPCC (2006) give estimates for aboveground biomass in agriculture land with perennial woody biomass (e.g., fruit orchards, agroforestry, etc.). Tables 4.7, 4.8, and 4.12 give aboveground biomass estimates for natural and plantation forest types. Recently, Ruesch and Gibbs (2008) mapped the IPCC (2006) aboveground biomass carbon storage data given year 2000 land cover data.

Other general sources of carbon storage estimates can be found. For example, Grace et al. (2006) estimate the average aboveground carbon storage (leaf + wood) for major savanna ecosystems around the world (Table 1). Houghton (2005) gives aboveground carbon storage for natural and plantation forest types, by continent (Tables 1 and 3). Brown et al. (1989) give aboveground biomass estimatesfor tropical broadleaf forests as a function of land-use: undisturbed, logged, nonproductive (Table 7). Region-specific sources of carbon storage data are also available. Those we've found include:

- Latin America: Malhi et al. (2006) report aboveground biomass volumes for 227 lowland forest plots in Bolivia, Brazil, Colombia, Ecuador, French Guinea, Guyana, Panama, Peru, and Venezuela. Nascimento and Laurance (2002) estimate aboveground carbon stocks in twenty 1-ha plots of Amazonian rainforest. Tiessen et al. (1998) find aboveground carbon stocks for the Brazilian savanna types Caatingas and Cerrados.
- Africa: Zhang and Justice (2001) report aboveground carbon stocks for major forest and shrub LULC types for central African countries. Tiessen et al. (1998) estimates total aboveground biomass of degraded savanna in Senegal. Makundi (2001) reports mean annual incremental growth for three forest plantation types in Tanzania. Malimbwi et al. (1994) estimates aboveground carbon stocks in the miombo woodlands of Kitungalo Forest Reserve Tanzania. Munishi and Shear (2004) report aboveground carbon stocks in the Afromontane rain forests

of the Eastern Arc Mountains of Tanzania. Glenday (2006) estimates aboveground carbon stocks for 3 forest types in the Kakamega National Forest of western Kenya.

- North America: Smith et al. (2006) estimate aboveground carbon stocks for all major forest types in the US.
- The Carbon On Line Estimator (http://www.ncasi2.org/COLE/) is a tool for calculating carbon characteristics in U.S. forests based on USDA Forest Service Forest Inventory & Analysis and Resource Planning Assessment data. With this tool, carbon characteristics can be examined at the scale of counties. Using the variables tab, aboveground, belowground, soil, or dead wood carbon pools can be selected.
- Other: Coomes et al. (2002) estimate aboveground carbon stocks for native shrubland and forest types in New Zealand.

One can also calculate aboveground biomass (and therefore carbon stocks) from timber inventories, which are often done by forestry ministries on a set of plots. Use the following formula to estimate the aboveground carbon stock in a forest stand that has been inventoried for its merchantable volume, where VOB is the per-hectare volume of trees in cubic meters measured from tree stump to crown point (the merchantable portion of the tree), WD is the wood density of trees (dry biomass per unit of tree volume), BEF is the ratio of total aboveground dry biomass to dry biomass of inventoried volume, and CF is the ratio of elemental carbon to dry biomass, by mass (Brown 1997). The biomass expansion factor (BEF) accounts for C stored in all other portions of the tree aboveground (e.g., branches, bark, stems, foliage, etc; the non-merchantable portions of the tree). In most cases WD for a plot is approximated with values for dominant species. Brown (1997) provides a table of WD values for many tree species in Appendix 1 of section 3 and a method for calculating BEF (Equation 3.1.4). See ECCM (2007) for an application of this FAO method to forest inventory data from eastern Tanzania. IPCC (2006) also presents estimates of ( ) where BEF values for hardwood, pine, conifer, and natural forest stands by eco-region are given in Table 4.5 and WD values for many species are given in Tables 4.13 and 4.14. (Use the BCEF values in Table 4.5 that are subscripted by S.) Finally, Brown et al. (1989) give BEF for tropical broadleaf forests under three land uses: undisturbed, logged, and nonproductive.

Brown (1997) attaches several caveats to the use of the above equation. First, the equation is designed for inventoried stands that are closed as opposed to open (forests with sparser canopy coverage such as oak savanna). Second, VOB estimates should be a function of all tree species found in the stand, not just the economically most valuable wood. Third, trees with diameters as low as 10 centimeters at breast height (DBH = 10) need to be included in the inventory if this aboveground biomass carbon equation is to be as accurate as possible. Brown (2002) also notes that the use of a single BEF value is a simplification of the actual biomass growth process.

These caveats lead Brown (2002) to recommend the use of allometric biomass equations to estimate woody aboveground biomass if available. These equations give the estimated relationship between a stand's distribution of different-sized trees and the stand's aboveground biomass. Brown (1997) and Brown and Schroeder (1999) provide general aboveground biomass allometric equations for all global eco-regions and the eastern US, respectively. Cairns et al. (2000) provide aboveground biomass allometric equations for LULC types in southern Mexico. Nascimento and Laurance (2002) estimate Amazonian rainforest aboveground biomass using allometric curves. The use of these equations requires knowledge of the distribution of tree size in a given stand.

Some researchers have made use of these equations a bit easier by first relating a stand's distribution of different-sized trees to its age and then mapping the relationship between age and aboveground biomass (i.e.,). For example, Silver et al. (2000) have estimated aboveground biomass as a function of stand age (i.e., years since afforestation/ reforestation) or previous LULC for native forest types in tropical ecosystems. Smith et al. (2006) take the transformation of allometric equations one step further by relating age to total biomass carbon (belowground plus aboveground) directly for various US forests.

When using IPCC data or other similar broad data sources, one final issue to consider is how the level of anthropogenic disturbance affects carbon stocks. The aboveground C stock of highly disturbed areas will likely be lower than the stocks of undisturbed areas. It is not clear what type of disturbance levels IPCC or other such sources assume when reporting aboveground biomass estimates. If forest disturbance is an issue in the demonstration site, LULC types should be stratified by levels of disturbance. For an example of such stratification see Table 2.5, page 14 of ECCM (2007). The effect of this disturbance on C storage in harvested wood products (HWPs) is discussed below.

Finally, we generally do nottreat aboveground herbaceous material as a carbon pool (e.g., grass, flowers, non-woody crops). Our working assumption is that this material does not represent a potential source of long-term storage like woody biomass, belowground biomass, and soil. Herbaceous material in general recycles its carbon too quickly.

2.2. Carbon stored in belowground biomass

For LULC categories dominated by woody biomass, belowground biomass can be estimated roughly with the "root to shoot" ratio of belowground to aboveground biomass. Default estimates of the root to shoot ratio are given in Table 4.4 on p. 4.49 of IPCC (2006) by eco-region. Broad estimates of this ratio are also given in Section 3.5 of Brown (1997).

Some LULC types contain little to no woody biomass but substantial belowground carbon stocks (e.g., natural grasslands, managed grasslands, steppes, and scrub/ shrub areas). In these cases the root to shoot ratio described above does not apply. Belowground estimates for these LULC types are best estimated locally, but if local data are not available some global estimates can be used. The IPCC (2006) lists total biomass (aboveground plus belowground) and aboveground biomass for each climate zone in table 6.4 (p. 6.27). The difference between these numbers is a crude estimate of belowground biomass. . Recently, Ruesch and Gibbs (2008) mapped the IPCC (2006) aboveground biomass carbon storage data given year 2000 land cover data.

Several studies have compiled estimates of belowground biomass or root-to-shoot ratios for different habitat types. Among those we found:

- Grace et al. (2006) estimate the total average woody and herbaceous root biomass for major savanna ecosystems around the world (Table 1). Baer et al. (2002) and Tilman et al. (2006) estimate the C stored in the roots of plots restored to native C4 grasses in Nebraska and Minnesota, U.S. respectively, as a function of years since restoration (see Table 2 in Baer et al. (2002) and Figure 1D in Tilman et al. (2006)).
- Cairns et al. (1997) survey root-to-shoot ratios for LULC types across the world. Munishi and Shear (2004) use a ratio of 0.22 for Afromontane forests in the Eastern Arc forests of Tanzania. Malimbwi et al. (1994) use 0.20 for miombo woodlands in the same area of Tanzania. Coomes et al. (2002) use 0.25 for shrublands in New Zealand. Gaston et al. (1998) report a root-to-shoot ratio of 1 for African grass / shrub savannas.

2.3. Carbon stored in soil

If local or regional soil C estimates are not available, default estimates can be looked up from IPCC (2006) for agricultural, pasture, and managed grasslands. Table 2.3 of IPCC (2006) contains estimates of soil carbon stocks by soil type, assuming these stocks are at equilibrium and have no active land management. For cropland and grasslandLULC types, this default estimate can be multiplied by management factors, listed in Tables 5.5 and 6.2 of IPCC (2006). For all other LULC types and their related management schemes, the IPCC (2006) assumes no management factors.

There are alternative global-level sources of soil carbon data. Post et al. (1982) report carbon stocks in the first meter of soil by Holdridge Life Zone Classification System (GIS map of these Zones available at http://www.arcgis.com/home/item.html?id=f3ec7241777f4c56a69ae14d2a98e44b). Silver etal. (2000) have estimated soil carbonas a function of years since afforestation / reforestation for native forest types in tropical ecosystems. Grace et al. (2006) estimate the soil carbon for major savanna types around the world (Table 1). Detwiler (1986) lists soil carbon for tropical forest soils in Table 2.

Several region-specific studies also report soil carbon stocks. Those we've found include:

- North America: Smith et al. (2006) estimate soil C for every 5-year increment up to 125 years since afforestation/ reforestation for all major forest types and forest management practices in each region of the U.S. Others include McLauchlan et al. (2006); Tilman et al. (2006); Fargione et al (2008); Schuman et al. (2002); and Lal (2002).
- Africa: Houghton and Hackler (2006) give soil C for 5 LULC forest types (Rain Forest; Moist Forest Dry; Forest; Shrubland; and Montane Forest) in sub-Saharan Africa that have retained their natural cover and for forest areas that have been converted to croplands, shifting cultivation, and pasture. Vagen et al. (2005) provides soil C estimates for various LULC types in sub-Saharan Africa.

- South America: Bernoux et al. (2002) estimated soil C stocks to a depth of 30 cm for different soil type-vegetation associations in Brazil. For example, the soil C stock in HAC soils under 14 different land cover categories, including Amazon forest and Brazilian Cerrado, range from 2 to 116 kg C m-2.

Important Note: In most research that estimates carbon storage and sequestration rates on a landscape, soil pool measures only include soil organic carbon (SOC) in mineral soils (Post and Kwon 2000). However, if the ecosystem being modeled has a lot of organic soils (e.g. wetlands or paramo), it is critical to add this component to the mineral soil content. In landscapes where the conversion of wetlands into other land uses is common, carbon releases from organic soils should also be tracked closely (IPCC 2006).

2.4. Carbon stored in dead organic matter

If local or regional estimates of carbon stored in dead organic matter aren't available, default values from the IPCC (2006) can be assigned. Table 2.2 (p. 2.27) gives default carbon stocks for leaf litter in forested LULC types. For non-forested types, litter is close to 0. Grace et al. (2006) estimate the average carbon stored in litter for major savanna ecosystems around the world (Table 1). It is not clear if their total "above-ground biomass" estimates include deadwood or not. Deadwood stocks are more difficult to estimate in general, and we have located no default data sources.

Regional estimates:

- United States: Smith et al. (2006) estimate carbon storage in litter (referred to as "Forest Floor" C in the document) and dead wood (the aggregate of C pools referred to as "Standing Dead Trees" and "Down Dead Wood" in the document) for all major forest types and forest management practices in each region of the U.S. as a function of stand age.
- South America: Delaney et al. (1998) estimate carbon stored in standing and down dead wood in 6 tropical forests of Venezuela. According to the authors, deadwood is typically 1/10 the amount of biomass as above-ground vegetation.

3. Decay rates for harvested wood products

For more information on the decay of carbon in HWP and methods for estimating it, see Skog et al. (2004), Green et al. (2006), Miner (2006), Smith et al. (2006), chapter 12, "Harvested Wood Products," of IPCC (2006), and Dias et al. (2007).

4. Harvest rates and dates harvest began

For an example of estimating carbon content in harvested wood products, we can use data from Makundi (2001). Assume that a softwood plantation in Tanzania has been producing timber for 50 years on a 5-hectare plot. Further, the rotation period for this type of plantation is 25 years (Makundi 2001). Assume an even age forestry operation. Therefore, every year, 2 hectares with 25-year old trees are clear-cut. The mean annual increment of the softwood's aboveground biomass is 17.82 Mg ha^{-1} yr^{-1} (Makundi 2001). Thus 2 hectares x 25 years x 17.82 Mg ha^{-1} yr^{-1}= 891 Mg of timber has been removedfrom the plantation annually for 50 years. If we assume the carbon content of the plantation's trees are 0.48 (Makundi 2001) then 891 x 0.48 = 427.68 metric tons of C are in the aboveground biomass of forest stand removed each year from the plantation or 8.6 $ha^{-1}yr^{-1}$.

Ascertaining dates in which harvesting began in each parcel may be difficult. If it is, you could assign an early date of initial harvest to all parcels, which essentially assumes that the carbon in the pool of harvested wood products has reached steady state (i.e., does not change year to year). Assume a date such that the time since first harvest is more than twice the half-life of carbon in the harvested wood products (e.g., if the half life of carbon in wood products is 20 years, choose a date of initial harvest that is 40 years before the current landscape map used.

5. Economic inputs: carbon price and discount rates

Recent estimates suggest that the social cost of carbon (SCC), or the marginal damage associated with the release of an additional Mg of C into the atmosphere, ranges from $32 per metric ton of C (Nordhaus 2007a) to $326 per metric ton of C (Stern 2007) in 2010 US dollars. The value of this damage can also be considered the monetary benefit of an avoided release. Tol (2009) provides a comprehensive survey of SCC estimates, reporting median values of $66 and $130 per metric ton in 2010 US dollars (values differ because of different assumptions regarding discounting of time). Other recent estimates can be found in Murphy et al. (2004), Stainforth et al. (2005), and Hope (2006).

An alternative method for measuring the cost of an emission of a metric ton of C is to set the cost equal to the least cost alternative for sequestering that ton. The next best alternative currently is to capture and store the C emitted from utility plants. According to Socolow (2005) and Socolow and Pacala (2007), the cost of this technology per metric ton captured and stored is approximately $100.

Finally, while we do not recommend this approach, market prices can be used to set the price of sequestered carbon. The Chicago Climate Exchange (CCX) and the European Climate Exchange (ECX) provide values ($24 and $153 per metric ton of C on May 14, 2008, respectively). The difference in these prices illustrates the problem with using markets to set values. The CCX and ECX are different in structure, scope, and the public policy that grounds each institution. This leads to different market fundamentals, and different prices for reasons unrelated to the social value of carbon sequestration. We do not recommend the use of market prices because they usually only apply to "additional" carbon sequestration; sequestration above and beyond some baseline sequestration rate. Further, carbon credit values from carbon markets such as the Chicago or European Climate Exchanges are largely a function of various carbon credit market rules and regulations and do not necessarily reflect the benefit to society of a sequestered ton of carbon. Therefore, correct use of market prices would require estimating a baseline rate for the landscape of interest, mapping additional sequestration, and then determining which additional sequestration is eligible for credits according to market rules and regulations. If the user is specifically interested in such an analysis please contact the InVEST team on the forums at http://ncp-yamato.stanford.edu/natcapforums/

We discount the value of future payments for carbon sequestration to reflect society's preference for payments that occur earlier rather than later. The US Office of Management and Budget recommends a 7% per annum market discount rate for US-based projects (OMB 1992). Discount rates vary for other parts of the world. Canada and New Zealand recommend 10% for their projects (Abusah and de Bruyn 2007).

Some economists believe that a market or consumption discount rate of 7% to 12% is too high when dealing with the climate change analysis. Because climate change has the potential to severely disrupt economies in the future, the preference of society to consume today at the expense of both climate stability in the future and future generations' economic opportunities is seen as unethical by some (Cline 1992, Stern 2007). According to this argument, analyses of the effects of climate change on society and policies designed to reduce climate change should use low discount rates to encourage greater GHG emission mitigation and therefore compensate for the potentially severe damages incurred by future generations (e.g., $r = 0.014$ in Stern (2007)). Recent government policies in several countries have supported the use of a very low discount rate for certain long-term projects (Abusah and de Bruyn 2007).

The carbon discount rate, which reflects the greater climatic impact of carbon sequestered immediately over carbon sequestered in the future, is discussed in Adams et al. (1999), Plantinga et al. (1999), Feng 2005, and Nelson et al. (2008).

5.2.8 References

Abusah, Sam and Bruyn, Clinton de. 2007. Getting Auckland on Track: Public Transport and New Zealand's Economic. Ministry of Economic Development Working Paper. Accessed at <http://s3.amazonaws.com/zanran_storage/www.med.govt.nz/ContentPages/4013253.pdf>.

Adams, DM, RJ Alig, BA McCarl, et al. 1999. Minimum cost strategies for sequestering carbon in forests. Land Econ75: 360-374.

Anderson, JR, EE Hardy, JT Roach, RE Witmer. A Land Use and Land Cover Classification System for Use with Remote Sensor Data. Washington, DC: United States Government Printing Office; 1976. Geological Survey Professional Paper 964.

Antle, JM, and B. Diagana. 2003. Creating Incentives for the Adoption of Sustainable Agricultural Practices in Developing Countries: The Role of Soil Carbon Sequestration. American Journal of Agricultural Economics85:1178-1184.

Baer, SG, DJ Kitchen, JM Blair, and CW Rice. 2002. Changes in Ecosystem Structure and Function along a Chronosequence of Restored Grasslands. Ecological Applications 12:1688-1701.

Bernoux, M., MDS Carvalho, B. Volkoff, and CC Cerri. 2002. Brazil's soil carbon stocks. Soil Science Society of America Journal66:888-896.

Brown, SL, PE Schroeder and JS Kern. Spatial distribution of biomass in forests of the eastern USA.Forest Ecology and Management 123 (1999: 81-90.

Brown, S. 2002. Measuring carbon in forests: current status and future challenges. Environmental Pollution116:363-372.

Brown, S. Estimating Biomass and Biomass Change of Tropical Forests: a Primer. FAO Forestry Department; 1997. Report for FAO Forestry Paper 134.

Brown, S. and PE Schroeder. 1999. Spatial patterns of aboveground production and mortality of woody biomass for eastern US forests. Ecological Applications9:968-980.

Cairns, MA, PK Haggerty, R. Alvarez, BHJ De Jong, and I. Olmsted. 2000. Tropical Mexico's recent land-use change: A region's contribution to the global carbon cycle. Ecological Applications 10:1426-1441.

Cairns, MA, S. Brown, EH Helmer, and GA Baumgardner. 1997. Root biomass allocation in the world's upland forests. Oecologia111:1-11.

Canadell, JG and MR Raupach. 2008. Managing Forests for Climate Change Mitigation. Science320:1456-1457.

Cline, WR. 1992. The economics of global warming. Instuitute for International Economics, Washington, D.C.

Coomes, DA, RB Allen, NA Scott, C. Goulding, and P. Beets. 2002. Designing systems to monitor carbon stocks in forests and shrublands. Forest Ecology and Management164:89-108.

Conte, MN and MJ Kotchen. 2010. Explaining the price of voluntary carbon offsets. Climate Change Economics 1 (2):93-111.

Capoor, K., and P. Ambrosi. State and Trends of the Carbon Market 2008. Washington, D.C.: World Bank Institute, 2008 May.

Delaney, M., S. Brown, AE Lugo, A. Torres-Lezama, and NB Quintero. 1998. The quantity and turnover of dead wood in permanent forest plots in six life zones of Venezuela. Biotropica30:2-11.

Detwiler, RP. 1986. Land Use Change and the Global Carbon Cycle: The Role of Tropical Soils. Biogeochemistry2:67-93.

Dias, AC, M. Louro, L. Arroja, and I. Capela. 2007. Carbon estimation in harvested wood products using a country-specific method: Portugal as a case study. Environmental Science & Policy 10 (3):250-259.

Edinburgh Centre for Carbon Management. The Establishing Mechanisms for Payments for Carbon Environmental Services in the Eastern Arc Mountains, Tanzania; 2007 May 2007.

Fargione, J., J. Hill, D. Tilman, S. Polasky, and P. Hawthorne. 2008. Land Clearing and the Biofuel Carbon Debt. Science319:1235-1238.

Feng, H. 2005. The dynamics of carbon sequestration and alternative carbon accounting, with an application to the upper Mississippi River Basin. Ecological Economics54:23-35.

Gaston, G., S. Brown, M. Lorenzini, and KD Singh. 1998. State and change in carbon pools in the forests of tropical Africa. Global Change Biology4:97-114.

Glenday, J. 2006. Carbon storage and emissions offset potential in an East African tropical rainforest. Forest Ecology and Management235:72-83.

Grace, J., J. San Jose, P. Meir, HS Miranda, and RA Montes. 2006. Productivity and carbon fluxes of tropical savannas. Journal of Biogeography33:387-400.

Green, C, V. Avitabile, EP Farrell, and KA Byrne. 2006. Reporting harvested wood products in national greenhouse gas inventories: Implications for Ireland. Biomass and Bioenergy 30(2): 105-114.

Gibbs, HK, S Brown, JO Niles, and JA Foley. 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. Environmental Research Letters2:045023.

Hamilton, K., M Sjardin, T Marcello, and G Xu. Forging a Frontier: State of the Voluntary Carbon Markets 2008. Washington, D.C.: Ecosystem Marketplace and New Carbon Finance; 2008.

Hope, CW. 2006. The social cost of carbon: what does it actually depend on? Climate Policy 6: 565–572

Houghton, RA. 2005. Tropical deforestation as a source of greenhouse gas emissions. In: Tropical Deforestation and Climate Change, Moutinho and Schwartzman [eds.]. Instituto de Pesquisa Ambiental da Amazonia and Environmental Defense, Belem,Brazil.

Houghton, RA, and JL Hackler. 2006. Emissions of carbon from land use change in sub-Saharan Africa. Journal of Geophysical Research111.

The Intergovernmental Panel on Climate Change (IPCC). 2006. 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 4: Agriculture, Forestry and Other Land Use. Prepared by the National Greenhouse Gas Inventories Programme, Eggleston, HS, L. Buendia, K. Miwa, T. Ngara, and K. Tanabe (eds). Institute for Global Environmental Strategies (IGES), Hayama, Japan. <http://www.ipcc-nggip.iges.or.jp/public/2006gl/vol4.html>.

Jenny, H. 1980. The Soil Resource. Springer, New York.

Lal, R. 2004. Soil Carbon Sequestration Impacts on Global Climate Change and Food Security. Science304:1623-1627.

Mackey, B, Keith H, Berry S.L, Lindenmayer DB. Green carbon: the role of natural forests in carbon storage. Part 1, A green carbon account of Australia's Southeastern Eucalypt forest, and policy implications. Canberra, Australia: ANU E Press, 2008.

Makundi, WR. 2001. Carbon mitigation potential and costs in the forest sector in Tanzania. Mitigation and Adaptation Strategies for Global Change 6:335-353.

Malhi, Y., D. Wood, TR Baker, et al. 2006. The regional variation of aboveground live biomass in old-growth Amazonian forests. Global Change Biology12:1107-1138.

Malimbwi, RE, B. Solberg, and E. Luoga. 1994. Estimation of biomass and volume in miombo woodland at Kitungalo Forest Reserve Tanzania. Journal of Tropical Forest Science7:230-242.

McLauchlan, KK., SE Hobbie, and WM Post. 2006. Conversion From Agriculture To Grassland Builds Soil Organic Matter On Decadal Timescales. Ecological Applications16:143-153.

Miner R. 2006. The 100-Year Method for Forecasting Carbon Sequestration in Forest Products in Use. Mitigation and Adaptation Strategies for Global Change (On-line only: http://www.springerlink.com/content/2l67274117366751/fulltext.pdf)

Mollicone D., F. Achard, S. Federici, H. Eva, G. Grassi, A. Belward, F. Raes, G. Seufert, H. Stibig, G. Matteucci, and E. Schulze. 2007. An incentive mechanism for reducing emissions from conversion of intact and non-intact forests. Climatic Change83:477-493.

Munishi, PKT and TH Shear. 2004. Carbon Storage in Afromontane Rain Forests of the Eastern Arc Mountains of Tanzania: their Net Contribution to Atmospheric Carbon. Journal of Tropical Forest Science16:78-93.

Murphy, JMet al. 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430, 768–772.

Murray, B., B. Sohngen, and M. Ross. 2007. Economic consequences of consideration of permanence, leakage and additionality for soil carbon sequestration projects. Climatic Change80:127-143.

Nascimento, HEM, and WF Laurance. 2002. Total aboveground biomass in central Amazonian rainforests: a landscape-scale study. Forest Ecology and Management168:311-321.

Nelson, E., G. Mendoza, J. Regetz, S. Polasky, H. Tallis, D. Cameron, K. Chan, G. Daily, J. Goldstein, P. Kareiva, E. Lonsdorf, R. Naidoo, TH Ricketts, and R. Shaw. 2008. Modeling Multiple Ecosystem Services and Tradeoffs at Landscape Scales. Frontiers in Ecology and the EnvironmentForthcoming.

Nordhaus, W. 2007a. Critical Assumptions in the Stern Review on Climate Change. Science 317 (5835): 201–202.

Nordhaus, W. 2007b. A Review of the Stern Review on the Economics of Global Warming. Journal of Economic Literature 45: 686-702.

Pagiola, S. 2008. Payments for environmental services in Costa Rica. Ecological Economics 65 (4): 712-724.

Plantinga, AJ, and RA Birdsey. 1994. Optimal Forest Stand Management When Benefits are Derived from Carbon. Natural Resource Modeling 8(4): 373-387.

Polasky, S, E Nelson, D Pennington, and K Johnson. 2010. The Impact of Land-Use Change on Ecosystem Services, Biodiversity and Returns to Landowners: A Case Study in the State of Minnesota. Environmental and Resource Economics, in press.

Post, WM, WR Emanuel, PJ Zinke, and AG Stangenberger. 1982. Soil carbon pools and world life zones. Nature298:156-159.

Post, WM, KC Kwon. 2000. Soil carbon sequestration and land-use change: processes and potential. Global Change Biology6:317-327.

Raich, JW, AE Russell, K. Kitayama, WJ Parton, and PM Vitousek. 2006. Temperature influences carbon accumulation in moist tropical forests. Ecology87:76-87.

Ruesch A, and HK Gibbs. 2008. New IPCC tier-1 global biomass carbon map for the year 2000. Available:http://cdiac.ornl.gov/epubs/ndp/global_carbon/carbon_documentation.html. Accessed 2008 Jul 7.

Schuman, GE, HH Janzen, and JE Herrick. 2002. Soil carbon dynamics and potential carbon sequestration by rangelands. Environmental Pollution, 116:391-396.

Sedjo, RA and B. Sohngen. Carbon Credits for Avoided Deforestation. Washington, DC: Resources for the Future; 2007 October 2007. Report for RFF DP 07-47.

Silver, WL, R. Ostertag, and AE Lugo. 2000. The potential for carbon sequestration through reforestation of abandoned tropical agricultural and pasture lands. Restoration Ecology8:394-407.

Skog, KE, K. Pingoud, and JE Smith. 2004. Method Countries Can Use to Estimate Changes in Carbon Stored in Harvested Wood Products and the Uncertainty of Such Estimates. Environmental Management 33, Supplement 1: S65–S73.

Smith, JE, LS Heath, KE Skog, RA Birdsey. Methods for Calculating Forest Ecosystem and Harvested Carbon with Standard Estimates for Forest Types of the United States. Newtown Square, PA: US Department of Agriculture, Forest Service, Northeastern Research Station; 2006. Report for NE-343.

Socolow, RH. 2005. Can We Bury Global Warming? Scientific American 293: 49-55.

Socolow, RH and SW Pacala. 2006. A Plan to Keep Carbon in Check. Scientific American 295: 50-57.

Sohngen, Brent, RH Beach, and Kenneth Andrasko. 2008. Avoided Deforestation as a Greenhouse Gas Mitigation Tool: Economic Issues. Journal of Environmental Quality 37: 1368-1375.

Stainforth, DA et al., 2005. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature 433, 403–406.

Stern, N. 2007. The Economics of Climate Change: The Stern Review. Cambridge and New York: Cambridge University Press.

Tiessen, H., C. Feller, EVSB Sampaio, and P. Garin. 1998. Carbon Sequestration and Turnover in Semiarid Savannas and Dry Forest. Climatic Change40:105-117.

Tilman, D., J. Hill, and C. Lehman. 2006. Carbon-Negative Biofuels from Low-Input High-Diversity Grassland Biomass. Science314:1598-1600.

Tol, RSJ. 2005. The marginal damage costs of carbon dioxide emissions: an assessment of the uncertainties. Energy Policy33:2064-2074.

Tol, RSJ. 2009. The Economic Effects of Climate Change.Journal of Economic Perspectives23: 29–51.

USOMB (US Office of Management and Budget). 1992. Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs Circular No. A-94 (Revised). Transmittal Memo No. 64. Washington DC: US Office of Management and Budget.

Vagen, TG, R Lal, and BR Singh. 2005. Soil carbon sequestration in sub-Saharan Africa: A review. Land Degradation & Development16:53-71.

Weitzman, ML. 2007. A review of the Stern Review on the Economics of Climate Change. Journal of Economic Literature45:703-724.

Zhang, Q, and CO Justice. 2001. Carbon Emissions and Sequestration Potential of Central African Ecosystems. AMBIO30:351-355.

5.3 Coastal Blue Carbon

5.3.1 Summary

Marine and terrestrial ecosystems help regulate Earth's climate by adding and removing greenhouse gases (GHGs) such as carbon dioxide (CO_2) to and from the atmosphere. Coastal marshes, mangroves and seagrasses, in particular, store large amounts of carbon in their sediments, leaves, and other forms of biomass. In addition to storing carbon, marine ecosystems continually accumulate carbon in their sediments, creating large reservoirs of long-term sequestered carbon. By storing and sequestereing carbon, marine ecosystems keep CO_2 out of the atmosphere where it would otherwise contribute to climate change.

Management activities that change the cover of coastal vegetation, such as the restoration of seagrass beds and the clearing of mangrove forests, change the ability of coastal and marine areas to store and sequester carbon.

The InVEST Coastal Blue Carbon model attempts to predict the amount of carbon stored and sequestered over a coastal zone at particular points in time due to changes in land cover. Using an estimate of the monetary social value, or where available, a market price for stored and sequestered carbon, the InVEST Coastal Blue Carbon model also quantifies the marginal value of storage and sequestration.

Results of the InVEST Coastal Blue Carbon model can be used to compare current and future scenarios of carbon stock and net sequestration, as well as identify locations within the landscape where degradation of coastal ecosystems should be avoided and restoration of coastal ecosystems should be prioritized in order to preserve and enhance these carbon storage and sequestration services.

5.3.2 Introduction

This model makes use of a variety of information, including:

- The distribution and abundance of coastal vegetation
- Habitat-specific carbon stock data
- Impact characteristics of various land-cover disturbances to biomass and soil carbon stock pools to predict carbon emission rates
- Carbon accumulation rates to estimate carbon stock, net sequestration and value across a land or seascape
- Estimates of the monetary social value or market price of carbon

To quantify the value of carbon storage and sequestration, the model focuses on changes in atmospheric carbon dioxide and other greenhouse gases as a result of changes caused by human activities that can affect marine ecosystems which store and sequester carbon. Changes to the composition of the atmosphere have wide-ranging effects on natural systems that can result in changes to agricultural productivity, air quality, sea levels, and more.

5.3.3 The Model

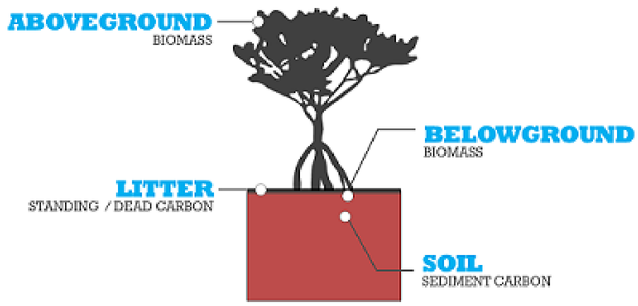
Modeling Considerations

Mapping and modeling changes in carbon storage and sequestration for coastal and marine habitats can present challenges. The types of spatial inputs and available information about the carbon cycle vary by location. Some study areas have high-quality data available for a detailed analysis while other locations do not have the information necessary to model changes in the position and function of coastal vegetation. Salt marsh, for example, is often studied in the context of migration due to sea-level rise. The combination of natural (e.g. sea-level rise) and anthropogenic (e.g. saltmarsh migration blocked by roads) factors should be included in scenario maps and subsequent carbon modeling where possible. When exploring future land cover scenarios, land cover map outputs produced by the SLAMM model (Sea Level Affecting Marshes Model, developed by Warren Pinnacle) can be useful inputs to the InVEST Coastal Blue Carbon model (Clougheet et al. 2010). However, because not all sites have the detailed elevation and habitat information required to run SLAMM, this InVEST model provides a flexible approach that allows users to provide either detailed land use/land cover maps or maps indicating the presence of coastal and marine vegetation that can sequester carbon.

How it Works

InVEST Coastal Blue Carbon models the carbon cycle through a bookkeeping-type approach (Houghton, 2003). This approach simplifies the carbon cycle by accounting for storage in three main pools (biomass, sediment carbon (i.e. soil), and standing dead carbon (i.e. litter) see Figure 1). Accumulation of carbon in coastal habitats occurs primarily in sediments (Pendleton et al., 2012). The model requires users to provide maps of coastal ecosystems that store carbon, such as mangroves and seagrasses. Users must also provide data on the amount of carbon stored in the three carbon pools and the rate of annual carbon accumulation in the biomass and sediments. If local information is not available, users can draw upon the global database of values for carbon stocks and accumulation rates sourced from the peer-reviewed literature that is included in the model. If data from field studies or other local sources are available, these values should be used instead of those in the global database. The model requires land cover maps, which represent changes in human use patterns in coastal areas or changes to sea level, to estimate the amount of carbon lost or gained over a specified period of time. The model quantifies carbon storage across the land or seascape by summing the carbon stored in these three carbon pools.

Figure 1. Three carbon pools for marine ecosystems included in the InVEST blue carbon model (mangrove example).



Carbon Storage

The carbon stored in a grid cell x at time t, given by S_{xt} and measured in Megatonnes of CO_2 equivalent, is equal to the sum of the carbon stored in each pool in the grid cell at any time (t),

$$S_{total} = S_{biomass} + S_{soil} + S_{litter}$$

where $S_{biomass}$, S_{soil}, S_{litter} indicate the respective Megatonnes of CO_2 equivalent stored in the biomass, soil and litter pools in a grid cell of a particular coastal blue carbon habitat.

Coastal blue carbon habitats can simply indicate the dominant vegetation type (e.g., eelgrass, mangrove, etc), or they can be based on details that affect pool storage values such as plant species, vegetation density, temperature regime, or vegetation age (e.g., time since restoration or last major disturbance).

For the sake of the carbon storage estimation, each coastal blue carbon habitat is assumed to be in storage equilibrium at any point in time (accumulation of carbon will be accounted for in the sequestration component of the model).

Carbon Accumulation

We model accumulation as the rate of carbon retained in the soil in organic form after the first year of decomposition. In relation to the annual ecosystem budget, this pool has not been remineralized, so it represents net accumulation. This carbon is usually derived from belowground production, and residence time can range from decades to millennia (Romero et al. 1994, Mateo et al. 1997). This accumulation contributes to the development of carbon "reservoirs" which are considered virtually permanent unless disturbed. Thus, even in the absence of a land-use or land-cover change, carbon continues to be sequestered naturally.

We estimate accumulation by multiplying habitat specific rates of carbon accumulation by the given cell area. The carbon accumulated in a grid cell x at time t in carbon pool p, given by A_{pt} and measured in Megatonnes of CO_2 equivalent per year, is equal to the rate of carbon accumulation in the sediments at time t.

Loss of carbon from the soil pool (sediments) upon disturbance is more nuanced than sequestration because different types of human uses and/or stasis may cause varied disruption of the soils and the carbon stored below. For example, high impact activities such as the clearing of mangroves for a shrimp pond or sediment dredging may result in a larger soil carbon disturbance than other activities such as commercial fishing or oil exploration. The impacts from coastal development on carbon storage varies since some types of development may involve paving over the soil, which often keeps a large percentage of the carbon stored intact. Alternatively, dredging could remove seagrasses and disturb the sediments below, releasing carbon into the atmosphere.

Carbon Emissions

When coastal ecosystems are degraded by human activities, the carbon stored in the living plant material (above and belowground) and the soil may be emitted to the atmosphere. The magnitude of post-conversion CO_2 release depends on the type of vegetation disturbed and the level of disturbance. The type of disturbance will determine the amount of aboveground biomass loss and depth to which the soil profile will be altered. The deeper the effects of the disturbance, the more soil carbon that will be exposed to oxygen, oxidized and consequently emitted in the form of CO_2. Some disturbances will only disturb the top soil layers while the deeper layers remain inundated and their carbon intact. Other disturbances may affect several meters of the soil profile. To estimate the extent of impact of various disturbances, we classify disturbances into three categories of impact: high, medium and low. Examples of high impact disturbances include mangrove conversion to shrimp farms and draining or diking salt marshes for conversion to agriculture. Low impact disturbance examples include recreational boating or float home marinas.

Magnitude and Timing of Loss We model the release of carbon from the biomass and soil pools by estimating the fraction of carbon lost from each pool's total stock at the time of disturbance. The fraction of carbon lost is determined by the original coastal blue carbon habitat and the level of impact resulting from the disturbance (see Table 1).

The InVEST Coastal Blue Carbon model allows users to provide details on the level of disturbance that occurs during a transition from a coastal blue carbon habitat to a non-coastal blue carbon habitat. This information can be provided to the model through a pre-processor tool (See "Transition Storage" section) and further clarified with an input transition table.

In general, carbon stock pools emit carbon at different rates: most emissions from the biomass pool take place within the first year, whereas emissions from the soil pool may take much longer. The model assigns exponential decay functions and half-live values to the biomass and soil carbon pools of each habitat type (Table 1; Murray et al. 2011).

Carbon emitted at time t due to a disturbance:

$$E_p = D_p \cdot (0.5^{\frac{t-(r+1)}{H_p}} - 0.5^{\frac{t-r}{H_p}})$$

where p is the carbon pool (biomass and soil), D_p represents the amount of carbon stock disturbed as time approaches infinity, H_p represents the half-life of the disturbance event, and E_p represents the share of carbon released from the total disturbed carbon stock at time t.

Rank	Salt marshes	Mangroves	Seagrasses	Other vegetation
% carbon loss from biomass	LI / MI: 50% biomass loss (1) HI: 100% biomass loss (1)	LI / MI: 50% biomass loss (1) HI: 100% biomass loss (1)	LI / MI: 50% biomass loss (1) HI: 100% biomass loss (1)	Use literature / field data
% carbon loss from soil^	LI: 30% loss (1) MI / HI: 100% loss (3)	LI: 30% loss (1) MI: 50% loss (1) HI: 66% loss (up to 1.5 m depth) (1)	LI / MI: top 10% washes away, bottom 90% decomposes in place (2) HI: top 50% washes away, bottom 50% decomposes in place (2)	Use literature / field data
Rate of decay (over 25 years)	Biomass half-life: 6 months (2) Soil half-life: 7.5 yrs (2)	Biomass half-life: 15 years, but assume 75% is released immediately from burning (2) Soil half-life: 7.5 years (2)	Biomass half-life: 100 days (2) Soil half-life: 1 year (2)	Use literature / field data
Methane emissions	1.85 T CO_2/ha/yr (4)	0.4 T CO_2/ha/yr	Negligible	Use literature / field data

Table 1: Percent carbon loss and habitat-specific decay rates as a result of low (LI), medium (MI) and high (HI) impact activities disturbing salt marsh, mangrove and seagrass ecosystems. These default values can be adjusted by modifying the input CSV tables.

References (numbers in parentheses above):

1. Donato, D. C., Kauffman, J. B., Murdiyarso, D., Kurnianto, S., Stidham, M., & Kanninen, M. (2011). Mangroves among the most carbon-rich forests in the tropics. Nature Geoscience, 4(5), 293-297.
2. Murray, B. C., Pendleton, L., Jenkins, W. A., & Sifleet, S. (2011). Green payments for blue carbon: Economic incentives for protecting threatened coastal habitats. Nicholas Institute for Environmental Policy Solutions, Report NI, 11, 04.
3. Crooks, S., Herr, D., Tamelander, J., Laffoley, D., & Vandever, J. (2011). Mitigating climate change through restoration and management of coastal wetlands and near-shore marine ecosystems: challenges and opportunities. Environment Department Paper, 121, 2011-009.
4. Krithika, K., Purvaja, R., & Ramesh, R. (2008). Fluxes of methane and nitrous oxide from an Indian mangrove. Current Science (00113891), 94(2).

Valuation of Net Sequestered Carbon

For interest rate i, discount rate d, and time t:

- $price_t = (1+i)^t \cdot price_0$ (if interest rate provided, else lookup in price table)

- $V = \sum_{t=0}^{T-1} \frac{price_t}{(1+d)^t} \cdot N_t$

where N is the net sequestered carbon, V is the net present value of net sequestered carbon, and T is the analysis year or final snapshot year if no analysis year is provided.

Identifying LULC Transitions with the Preprocessor

The land use / land cover (LULC) maps provide snapshots of a changing landscape and are the inputs that drive carbon accumulation and emissions in the model. The user must first produce a set of coastal and marine habitat maps via a land change model (e.g., SLAMM), a scenario assessment tool, or manual GIS processing. The user must then input the LULC maps into the model in chronological order (s_0, s_1, ..., s_n).

The preprocessor tool compares LULC classes across the maps to identify the set of all LULC transitions that occur. The preprocessor tool then generates a transition matrix that indicates whether a transition occurs between two habitats (e.g. salt marsh to developed dry land) and whether carbon accumulates, is disturbed, or remains unchanged once that transition occurs.

Land Cover Transition Types:

- Other LULC Class ⇒ Coastal Blue Carbon Habitat (*Carbon Accumulation* in Succeeding Years of Transition Event Until Next Bounding Year)
- Coastal Blue Carbon Habitat ⇒ Coastal Blue Carbon Habitat (*Carbon Accumulation* in Succeeding Years of Transition Event Until Next Bounding Year)
- Coastal Blue Carbon Habitat ⇒ Other LULC Class (*Carbon Disturbance* in Succeeding Years of Transition Event Until End of Time Series Forecast)
- Other LULC Class ⇒ Other LULC Class (*No Carbon Change* in Succeeding Years of Transition Event Until Next Bounding Year)

This transition matrix produced by the coastal blue carbon preprocessor, and subsequently edited by the user, allows the model to identify where human activities and natural events disturb carbon stored by vegetation. If a transition from one LULC class to another does not occur during any of the time steps, the cell will be left blank. For cells in the matrix where transitions occur, the tool will populate a cell with 'accum' in the cases where a non-coastal blue carbon habitat transitions to a coastal blue carbon habitat or a coastal blue carbon habitat transitions to another coastal blue carbon habitat, 'disturb' in the case where a coastal blue carbon habitat transitions to a non-coastal blue carbon habitat, or 'NCC' (for no carbon change) in the case where a non-coastal blue carbon habitat transitions to another non-coastal blue carbon habitat. For example, if a salt marsh pixel in t_0 is converted to developed dry land in t_1 then the cell will be populated with 'disturb'. On the other hand, if a mangrove remains a mangrove over this same time period then this cell in the matrix will be populated with 'accum'. It is likely that a mangrove that remains a mangrove will accumulate carbon in its soil and biomass.

The user will then need to modify the 'disturb' cells with either 'low-impact-disturb', 'med-impact-disturb' or 'high-impact-disturb' depending on the level of disturbance that occurs as the transition occurs between LULC types. This gives the user more fine-grained control over emissions due to disturbance. For example, rather than provide only one development type in an LULC map, a user can separate out the type into two development types and update the transition matrix accordingly so that the model can more accurately quantify and map changes in carbon as a result of natural and anthropogenic factors. Similarly, different species of mangroves may accumulate soil carbon at different rates. If this information is known, it can improve the accuracy of the model to provide this species distinction (two different classes in the LULC input maps) and then the associated accumulation rates in the Carbon Pool Transient Values CSV table.

Model Math

Dimensions

- x, y: Position
- t: Timestep (*Years Ahead of Baseline Year*)
- s: Snapshot Year (*Year in which a Snapshot is Provided*)
- r: Transition Year (*Year in which a Transition Event Begins*)
- a: Analysis Year (*Final Year of the Time Series Forecast*)
- b: Bounding Year (*Year that Bounds a Transition Event*)
- p: Carbon Pool

Multidimensional Matrices

- $C_{s,x,y}$: LULC Map (*unitless*)
- $S_{b,p,x,y}$: Carbon Stock for Biomass and Soil Pools (*Megatonnes CO2e*)
- $L_{s,x,y}$: Litter Pool Carbon Stock (*Megatonnes CO2e*)
- $T_{b,x,y}$: Total Carbon Stock (*Megatonnes CO2e*)
- $Y_{r,p,x,y}$: Yearly Accumulated Carbon (*Megatonnes CO2e / Year*)
- $A_{t,p,x,y}$: Accumulated Carbon (*Megatonnes CO2e / Year*)
- $A_{r,p,x,y}$: Accumulated Carbon (*Megatonnes CO2e / Transition*)
- $D_{r,p,x,y}$: Carbon Stock Disturbed (*Megatonnes CO2e*)
- $H_{r,p,x,y}$: Disturbed Carbon Stock Emissions Half-life (*Years*)
- $E_{t,p,x,y}$: Emitted Carbon (*Megatonnes CO2e / Year*)
- $E_{r,p,x,y}$: Emitted Carbon (*Megatonnes CO2e / Transition*)
- $N_{t,x,y}$: Net Sequestered Carbon (*Megatonnes CO2e / Year*)
- $N_{r,x,y}$: Net Sequestered Carbon (*Megatonnes CO2e / Transition*)
- $V_{x,y}$: Net Present Value of Net Sequestered Carbon (*$ at Baseline Year*)

Initial Conditions

- $S_{0,p,x,y}, L_{0,x,y}, T_{0,x,y} \Leftarrow$ reclass($C_{0,x,y}$, cell_size, lulc_carbon_stock_initial_conditions)

Time Series Forecast

- $Y_{r,p,x,y}, D_{r,p,x,y}, H_{r,p,x,y}, L_{s,x,y} \Leftarrow$ reclass($C_{s,x,y}$, $S_{b,x,y}$ cell_size, lulc_carbon_stock_transient_conditions)
- $A_{t,p,x,y} \Leftarrow$ compute_timestep_accumulation($Y_{r,p,x,y}$, t)
- $E_{t,p,x,y} \Leftarrow$ compute_timestep_emissions($D_{r,p,x,y}$, $H_{r,p,x,y}$, t)
- $N_{t,p,x,y} \Leftarrow$ compute_timestep_net_sequestration($A_{t,p,x,y}$, $E_{t,p,x,y}$, t)

- $A_{r,p,x,y} \Leftarrow$ compute_transition_period_total_accumulation($Y_{r,p,x,y}$, r)
- $E_{r,p,x,y} \Leftarrow$ compute_transition_period_total_emissions($E_{t,p,x,y}$, r)
- $N_{r,p,x,y} \Leftarrow$ compute_transition_period_net_sequestration($A_{r,p,x,y}$, $E_{r,p,x,y}$, r)
- $S_{b,p,x,y} \Leftarrow$ compute_carbon_stock($S_{b,p,x,y}$, $N_{r,p,x,y}$, b)
- $T_{b,x,y} \Leftarrow$ compute_carbon_stock_with_litter($S_{b,p,x,y}$, $L_{s,x,y}$, b)
- $V_{x,y} \Leftarrow$ compute_net_present_value($N_{t,p,x,y}$, $price_t$, $discount\_rate$)

Time Series Forecast Functions in Detail

compute_timestep_accumulation($Y_{r,p,x,y}$, t)

- $A_{t,p,x,y} \Leftarrow Y_{r,p,x,y}$

compute_timestep_emissions($D_{r,p,x,y}$, $H_{r,p,x,y}$, t)

- $E_{t,p,x,y} \Leftarrow \sum_{r_{prev}} \left(D_{r,p,x,y} \cdot (0.5^{\frac{t-(r+1)}{H_{r,p,x,y}}} - 0.5^{\frac{t-r}{H_{r,p,x,y}}}) \right)$

compute_timestep_net_sequestration($A_{t,p,x,y}$, $E_{t,p,x,y}$, t)

- $N_{t,x,y} \Leftarrow A_{t,x,y} - E_{t,x,y}$

compute_transition_period_total_accumulation($Y_{r,p,x,y}$, r)

- $A_{r,p,x,y} \Leftarrow (t_{b\_next} - t_{b\_prev}) \cdot Y_{r,p,x,y}$

compute_transition_period_total_emissions($E_{t,p,x,y}$, r)

- $E_{r,p,x,y} \Leftarrow \sum_{t_{b\_prev}}^{t_{b\_next}} E_{t,p,x,y}$

compute_transition_period_net_sequestration($A_{r,p,x,y}$, $E_{r,p,x,y}$, r)

- $N_{r,p,x,y} \Leftarrow A_{r,p,x,y} - E_{r,p,x,y}$

compute_carbon_stock($S_{b,p,x,y}$, $N_{r,p,x,y}$, b)

- $S_{b_{next},p,x,y} \Leftarrow S_{b_{prev},p,x,y} + N_{r,p,x,y}$

compute_carbon_stock_with_litter($S_{b,p,x,y}$, $L_{s,x,y}$ b)

- $T_{b,x,y} \Leftarrow L_{s,x,y} + \sum_p S_{b,p,x,y}$

compute_net_present_value($N_{t,p,x,y}$, $price_t$, $discount\_rate$)

- $V_{x,y} \Leftarrow \sum_t \left(\frac{price_t}{(1+discount\_rate)^t} \cdot N_{t,x,y} \right)$

Results

- $T_{b,x,y}$: Total Carbon Stock (*Megatonnes CO2e per Hectare*)
- $A_{r,x,y}$: Carbon Accumulation (*Megatonnes CO2e per Hectare*)
- $E_{r,x,y}$: Carbon Emissions (*Megatonnes CO2e per Hectare*)
- $N_{r,x,y}$: Net Carbon Sequestration (*Megatonnes CO2e per Hectare*)
- $V_{x,y}$: Net Present Value at Baseline Year (*$ per Hectare*)

5.3.4 Limitations and Simplifications

In the absence of detailed knowledge on the carbon dynamics in coastal and marine systems, we take the simplest accounting approach and draw on published carbon stock datasets from neighboring coastlines. We use carbon estimates from the most extensive and up-to-date published global datasets of carbon storage and accumulation rates (e.g., Fourqurean et al. 2012 & Silfeet et al. 2011).

- We assume all storage and accumulation occurs in the biomass and soil pools.
- We ignore increases in stock and accumulation with growth and aging of habitats.
- We assume that carbon is stored and accumulated linearly through time between the current and future scenarios.
- We assume that, after a disturbance event occurs, the disturbed carbon is emitted over time at an exponential decay rate.
- We assume that some human activities that may degrade coastal ecosystems do not disturb carbon in the sediments.

5.3.5 Data Needs

Biophysical Inputs

The following are the data needs for the biophysical portion of the InVEST Coastal Blue Carbon model:

- **Land Use/Land Cover (LULC) maps**: Maps of initial (t_1) and future (t_2) LULC (e.g., developed dry land, shrimp aquaculture, mangrove forest, salt marsh, etc).
- **Years of provided LULC maps**: (t_1, t_2, ...), the model uses these years to determine length of time (number of years; (t_2 - t_1) of the analysis and multiplies this value by the user-specified accumulation rates (Megatonnes of CO_2 e/ha/yr). If the user is only interested in the standing stock of carbon at t_1, then this input is optional.
- **Carbon pool initial values by LULC class**: A collection of values of carbon storage in biomass (Megatonnes of CO_2 e/ha), soil (Megatonnes of CO_2 e/ha), and litter (tonnes of CO_2/ha) for each LULC class.
- **Transition matrix**: A table produced by the preprocessor tool that indicates either disturbance or accumulation of carbon based on preprogrammed logic for LULC transitions from t_n to t_{n+1}. Disturbance values must be modified by user.
- **Carbon pool transient values by LULC class**: A collection of values on the accumulation rate (Megatonnes of CO_2 e/ha-yr), percent disturbance and half-lives of carbon emitted over time within the biomass and soil pools of each LULC class.

5.3.6 Running the InVEST Model

Step 1. The Preprocessor

Inputs

1. **Workspace Folder**: The selected folder is used as the workspace where all intermediate and final output files will be written. If the selected folder does not exist, it will be created. If datasets already exist in the selected folder, they will be overwritten.
2. **Results Suffix (Optional)**: This text will be appended to the end of the yield function output folders to help seperate outputs from multiple runs. Please see the **'Interpreting Results'_** section for an example folder structure for outputs.

3. **LULC Lookup Table (CSV)**: A CSV table used to map LULC classes to their values in a raster, as well as to indicate whether or not the LULC class is a coastal blue carbon habitat.

lulc-class	code	is_coastal_blue_carbon_habitat
<str>	<int>	<bool>
...	...	...

4. **LULC Snapshots (Rasters)**: A set of GDAL-supported rasters representing the land/seascape at particular points in time. Provided in chronological order.

int	int
int	int

Outputs

Output Folder Structure

```
.
|-- outputs
    |-- transitions.csv
    |-- carbon_pool_initial_template.csv
    |-- carbon_pool_transient_template.csv
```

Outputs

1. **LULC Transition Effect on Carbon Emissions (CSV)**: This transition matrix indicates whether disturbance or accumulation occurs in a transition from one LULC class to another. If the cell is left blank, then no transition of that kind occurs between snapshots. The left-most column represents the source LULC class, and the top row represents the destination LULC class. Depending on the transition type, a cell will be pre-populated with one of the following: (empty), 'NCC' (for no carbon change), 'accum', 'disturb'. It is up to the user to edit the 'disturb' cells with the degree to which distance occurs due to the change. This is done by changing 'disturb' to either 'low-impact-disturb', 'med-impact-disturb', or 'high-impact-disturb'.

lulc-class	<lulc1>	<lulc2>	...
<lulc1>	<str>	<str>	...
<lulc2>	<str>	<str>	...
...	...	...	...

2. **Carbon Pool Initial Variables Table (CSV)**: The user must fill in the 'biomass', 'soil', and 'litter' columns with amount of carbon initially stored in each pool of a lulc-class in terms of Megatonnes CO_2 e/ hectare. See *Step 2. The Main Model* for more information.

lulc-class	biomass	soil	litter
<str>			
...			

3. **Carbon Accumulation/Disturbance Transient Variables Table (CSV)**: The user must fill in all columns besides the 'lulc-class' and 'pool' columns. See *Step 2. The Main Model* for more information.

lulc-class	pool	half-life	yearly_accumulation	low-impact-disturb	med-impact-disturb	high-impact-disturb
<lulc1>	biomass					
<lulc1>	soil					
<lulc2>	biomass					
<lulc2>	soil					
...	...					

Step 2. The Main Model

Inputs

Workspace Folder: The selected folder is used as the workspace where all intermediate and final output files will be written. If the selected folder does not exist, it will be created. If datasets already exist in the selected folder, they will be overwritten.

Results Suffix (Optional): This text will be appended to the end of the yield function output folders to help separate outputs from multiple runs. Please see the **'Interpreting Results'_** section for an example folder structure of the outputs.

LULC Lookup Table (CSV): A CSV table used to map LULC classes to their values in a raster and to indicate whether or not the LULC class is a coastal blue carbon habitat.

lulc-class	code	is_coastal_blue_carbon_habitat
<str>	<int>	<bool>
...	...	...

LULC Snapshots (Rasters): A set of GDAL-supported rasters representing the landscape/seascape at particular points in time. Provided in chronological order.

LULC Snapshot Years: A set of years that respectively correspond to the provided LULC snapshot rasters. Provided in chronological order.

LULC Transition Effect on Carbon Emissions (CSV): Generated by the preprocessor. This file must be edited before it can be used by the main model. The left-most column represents the source LULC class, and the top row represents the destination LULC class.

lulc-class	<lulc1>	<lulc2>	...
<lulc1>	<str>	<str>	...
<lulc2>	<str>	<str>	...
...	...	...	...

Carbon Pool Initial Variables Table (CSV): The provided CSV table contains information related to the initial conditions of the carbon stock within each of the three pools of a habitat. Biomass includes carbon stored above and below ground. All non-coastal blue carbon habitat LULC classes are assumed to contain no carbon. The values for 'biomass', 'soil', and 'litter' should be given in terms of Megatonnes CO_2 e/ ha.

lulc-class	biomass	soil	litter
<str>	<float>	<float>	<float>
...	...	...	...

Carbon Accumulation/Disturbance Transient Variables Table (CSV): The provided CSV table contains information related to the transition of carbon into and out of coastal blue carbon pools. All non-coastal blue carbon habitat LULC classes are assumed to neither sequester nor emit carbon as a result of change. The 'yearly_accumulation' values should be given in terms of Megatonnes of CO_2 e/ha-yr. The 'half-life' values must be given in terms of years. The 'disturbance' values must be given as a decimal percentage of stock distrubed given a transition occurs away from a lulc-class.

lulc-class	pool	yearly_accumulation	half-life	low-impact-disturb	med-impact-disturb	high-impact-disturb
<lulc1>	biomass	<float>	<float>	<float>	<float>	<float>
<lulc1>	soil	<float>	<float>	<float>	<float>	<float>
<lulc2>	biomass	<float>	<float>	<float>	<float>	<float>
<lulc2>	soil	<float>	<float>	<float>	<float>	<float>
...	...	...	...	...	...	...

Price: The price per Megatonne CO_2 e at the base year.

Interest Rate: The interest rate on the price per Megatonne CO_2 e, compounded yearly.

Price Table (CSV): Can be used in place of price and interest rate inputs. The provided CSV table contains the price per Megatonne CO_2 e sequestered for a given year, for all years from the original snapshot to the analysis year, if provided.

year	price
...	...

Discount Rate: The discount rate on future valuations of sequestered carbon, compounded yearly.

Outputs

Output Folder Structure

```
.
|-- outputs
    |-- carbon_stock_at_[year].tif
    |-- carbon_accumulation_between_[year]_and_[year].tif
    |-- carbon_emissions_between_[year]_and_[year].tif
    |-- net_carbon_sequestion_between_[year]_and_[year].tif
    |-- net_present_value.tif
```

Outputs

1. **Carbon Stock Rasters**

- Units: Megatonnes CO_2 e per Hectare

2. **Carbon Accumulation Rasters**

- Units: Megatonnes CO_2 e per Hectare

3. **Carbon Emissions Rasters**

- Units: Megatonnes CO_2 e per Hectare

4. **Net Carbon Sequestration Rasters**

- Units: Megatonnes CO_2 e per Hectare

5. **Net Present Value Raster**

- Units: (Currency of Provided Price Table) per Hectare

5.3.7 Example Use-Case

Freeport, Texas

Summary

Over the next 100 years, the US Gulf coast has been identified as susceptible to rising sea levels. The use of the InVEST blue carbon model serves to identify potential changes in the standing stock of carbon in coastal vegetation that sequester carbon. This approach in Freeport, TX was made possible with rich and resolute elevation and LULC data sets. We used a 10-meter DEM with sub-meter vertical accuracy to model marsh migration and loss over time as a result of sea level rise using Warren Pinnacle's SLAMM (Sea Level Affected Marsh Model). Outputs from SLAMM serve as inputs to the InVEST blue carbon model which permits the tool to map, measure and value carbon sequestration and emissions resulting from coastal land cover change over a 94-year period.

The Sea Level Affecting Marshes Model (SLAMM: http://www.warrenpinnacle.com/prof/SLAMM/) models changes in the distribution of 27 different coastal wetland habitat types in response to sea-level rise. The model relies on the relationship between tidal elevation and coastal wetland habitat type, coupled with information on slope, land use, erosion and accretion to predict changes or loss of habitat. SLAMM outputs future habitat maps for user-defined time steps and sea-level rise scenarios. These future habitat maps can be utilized with InVEST service models to evaluate resultant changes in ecosystem services under various sea-level rise scenarios (e.g. 1 meter SLR by 2100).

For example, SLAMM was used to quantify differences in carbon sequestration over a range of sea-level rise projections in Galveston Bay, Texas, USA. First, SLAMM was used to map changes in the distribution of coastal wetland habitat over time under different sea-level rise projections. Then, the InVEST blue carbon model was used to evaluate changes in carbon sequestration associated with predicted changes in habitat type. The 27 land-cover classes modeled by SLAMM were condensed into a subset relevant to carbon sequestration and converted from ASCII to raster format for use with InVEST. SLAMM results produced LULC maps of future alternative scenarios over 25-year time slices beginning in 2006 and ending in 2100. The following figure depicts 2006 LULC and a table of disaggregated land class types.

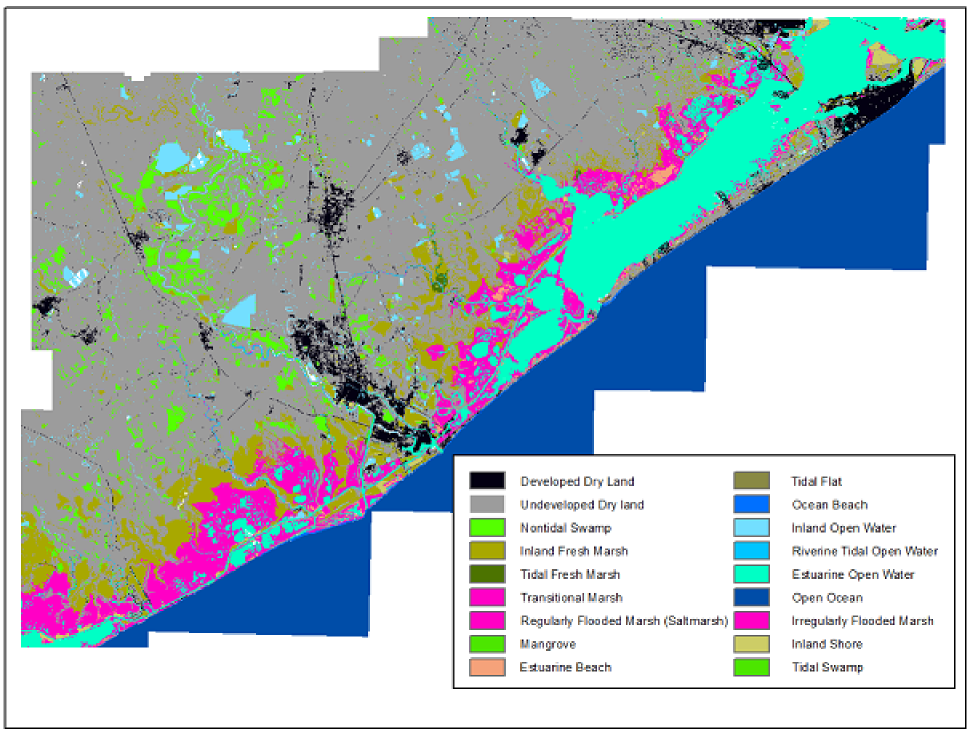


Figure CS1. Current (2006) LULC map of Freeport, Texas

Carbon stored in the sediment (‘soil’ pool) was the focus of this analysis. The vast majority of carbon is sequestered in this pool by coastal and marine vegetation. See the case study limitations for additional information. To produce maps of carbon storage at the different 25-year time steps, we used the model to perform a simple “look-up” to determine the amount of carbon per 10-by-10 meter pixel based on known storage rates from sampling in the Freeport area (Chmura et al. 2003).

Next, we provide the InVEST model with a transition matrix in order to identify the amount of carbon gained or lost over each 25-year tiume step. Annual accumulation rates in salt marsh were also obtained from Chmura et al. (2003). When analyzing the time period from 2025 to 2050, we assume t_2 = 2025 and t_3 = 2050. We identify all the possible transitions that will result in either accumulation or loss of carbon. The model compares the two LULC maps (t_2 and t_3) to identify any pixel transitions from one land cover type to another. We apply these transformations to the standing stock of carbon which is the running carbon tally at t_2 (2025). Once these adjustments are complete, we have a new map of standing carbon for t_3 (2050). We repeat this step for the next time period where t_3 = 2050 and t_4 = 2075. This process was repeated until 2100. The model produces spatially explicit depictions of net sequestration over time as well as summaries of net gain/emission of carbon for the two scenarios at each 25-year time step. This information was used to determine during which time period for each scenario the rising seas and resulting marsh migration led to net emissions for the study site and the entire Freeport area.

Time Period	Scenario #1: No Management	Scenario #2: High Green
2006-2025 (t_1-t_2)	+4,031,180	+4,172,370
2025-2050 (t_2-t_3)	-1,170,580	+684,276
2050-2075 (t_3-t_4)	-7,403,690	-5,525,100
2075-2100 (t_4-t_5)	-7,609,020	-8,663,600
100-Year Total:	-12,152,100	-9,332,050

Table CS1. Carbon sequestration and emissions for each 25-year time period for the two scenarios of the entire Freeport study area.

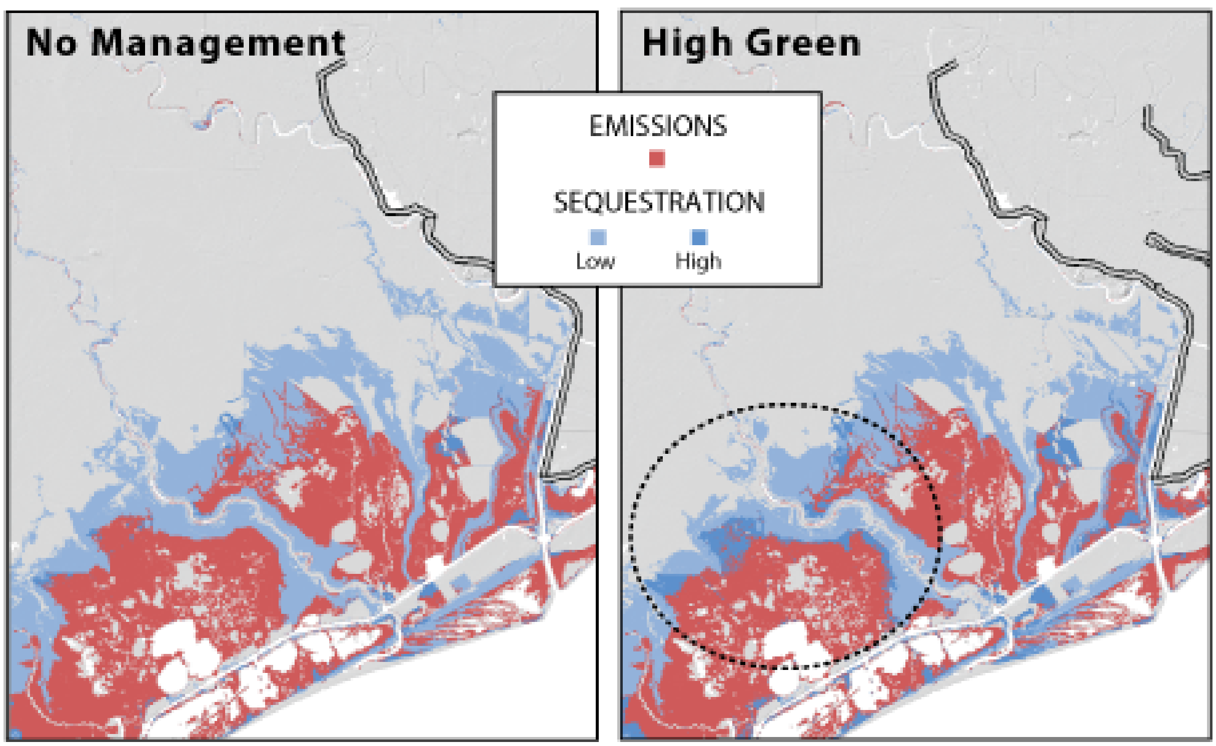


Figure CS2. Carbon emissions (red) and sequestration (blue) from 2006 to 2100 for the two scenarios and a subset of the Freeport study area.

The following is table summarizing how the main inputs, where they were obtained and how they were used in the model:

Input	Source	Use in the InVEST blue carbon model
DEM	USGS	DEM was needed to produce the future LULC maps using the SLAMM tool.
Land use / land cover (LULC)	USGS/NOAA	Salt marshes store carbon in biomass and soils. We utilized maps showing the current distribution of salt marshes to establish a baseline coverage of marshes from which we estimate aboveground biomass and soil carbon.
Carbon stock in salt marsh systems	Natural Capital Project literature review	Carbon storage was calculated by summing the carbon stored in biomass and sediments. Carbon stocks were calculated for all of the areas of functional salt marsh in the study region (Chmura et al. 2003).
Social value of carbon in 2006 US $	USIWGSCC 2010	The "social cost of carbon" (SCC) is an estimate of the monetized damages associated with an incremental increase in carbon emissions in a given year. It is intended to include (but is not limited to) changes in net agricultural productivity, human health, property damages from increased flood risk, and the value of ecosystem services. The social cost of carbon is useful for allowing institutions to incorporate the social benefits of reducing carbon dioxide (CO_2) emissions into cost benefit analyses of management actions that have small, or "marginal," impacts on cumulative global emissions.
Discount rate	USIWGSCC 2010	This discount rate reflects society's preferences for short run versus long term consumption. Since carbon dioxide emissions are long-lived, subsequent damages occur over many years. We use the discount rate to adjust the stream of future damages to its present value in the year when the emissions were changed.

Table CS2. Input summary table for using InVEST blue carbon model in Freeport, Texas

Limitations

- This analysis did not model change in carbon resulting from growth or loss of aboveground biomass of coastal and marine vegetation.
- While the spatial resolution of the LULC maps produced by SLAMM was very high (10 meters), the temporal resolution provided by SLAMM was quite coarse (25-year time steps). The carbon cycle is a dynamic process. By analyzing change over 25-year time periods, we ignore any changes that are not present at the start and end of each time step.

5.3.8 References

Baumol, W. J. (1968). On the social rate of discount. The American Economic Review, 788-802.

Bouillon, S., Borges, A. V., Castañeda-Moya, E., Diele, K., Dittmar, T., Duke, N. C., ... & Twilley, R. R. (2008). Mangrove production and carbon sinks: a revision of global budget estimates. Global Biogeochemical Cycles, 22(2).

Chmura, G. L., Anisfeld, S. C., Cahoon, D. R., & Lynch, J. C. (2003). Global carbon sequestration in tidal, saline wetland soils. Global biogeochemical cycles, 17(4).

Clough, J. S., Park, R., and Fuller, R. (2010). "SLAMM 6 beta Technical Documentation." Available at http://warrenpinnacle.com/prof/SLAMM.

Fourqurean, J. W., Duarte, C. M., Kennedy, H., Marbà, N., Holmer, M., Mateo, M. A., ... & Serrano, O. (2012). Seagrass ecosystems as a globally significant carbon stock. Nature Geoscience, 5(7), 505-509.

Hope, Chris. (2011) "The PAGE09 Integrated Assessment Model: A Technical Description." Cambridge Judge Business School Working Paper No. 4/2011 (April). Available at http://www.jbs.cam.ac.uk/research/working_papers/2011/wp1104.pdf.

Houghton, R. A. (2003). Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850–2000. Tellus B, 55(2), 378-390.

Pendleton, L., Donato, D. C., Murray, B. C., Crooks, S., Jenkins, W. A., Sifleet, S., ... & Baldera, A. (2012). Estimating global "blue carbon" emissions from conversion and degradation of vegetated coastal ecosystems. PLoS One, 7(9), e43542.

Rosenthal, A., Arkema, K., Verutes, G., Bood, N., Cantor, D., Fish, M., Griffin, R., and Panuncio, M. (In press). Identification and valuation of adaptation options in coastal-marine ecosystems: Test case from Placencia, Belize. Washington, DC: InterAmerican Development Bank. Technical Report.

Sifleet, S., Pendleton, L., and B. Murray. (2011). State of the Science on Coastal Blue Carbon. Nicholas Institute Report, 1-43.

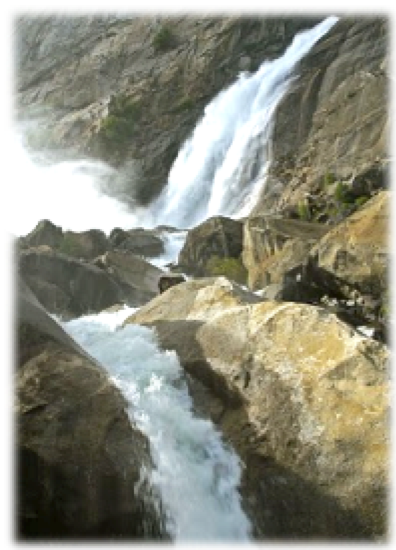
United States, Interagency Working Group on Social Costs of Carbon. (2010) "Technical Support Document: Social Cost of Carbon for Regulatory Impact Analysis Under Executive Order 12866." Available at http://www.epa.gov/otaq/climate/regulations/scc-tsd.pdf.

United States, Interagency Working Group on Social Costs of Carbon. (2013) "Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis Under Executive Order 12866." Available at http://www.whitehouse.gov/sites/default/files/omb/inforeg/social_cost_of_carbon_for_ria_2013_update.pdf.

World Bank. (2012). State and Trends of the Carbon Market 2012. Washington DC: The World Bank, 133.

5.4 Water Yield: Reservoir Hydropower Production

5.4.1 Summary



Hydropower accounts for twenty percent of worldwide energy production, most of which is generated by reservoir systems. InVEST estimates the annual average quantity and value of hydropower produced by reservoirs, and identifies how much water yield or value each part of the landscape contributes annually to hydropower production. The model has three components: water yield, water consumption, and hydropower valuation. The first two components use data on average annual precipitation, annual reference evapotranspiration and a correction factor for vegetation type, root restricting layer depth, plant available water content, land use and land cover, root depth, elevation, saturated hydraulic conductivity, and consumptive water use. The valuation model uses data on hydropower market value and production costs, the remaining lifetime of the reservoir, and a discount rate. The biophysical models do not consider surface – ground water interactions or the temporal dimension of water supply. The valuation model assumes that energy pricing is static over time.

5.4.2 Introduction

The provision of fresh water is an ecosystem service that contributes to the welfare of society in many ways, including through the production of hydropower, the most widely used form of renewable energy in the world. Most hydropower production comes from watershed-fed reservoir systems that generally deliver energy consistently and predictably. The systems are designed to account for annual variability in water volume, given the likely levels for a given watershed, but are vulnerable to extreme variation caused by land use and land cover (LULC) changes. LULC changes can alter hydrologic cycles, affecting patterns of evapotranspiration, infiltration and water retention, and changing the timing and volume of water that is available for hydropower production (World Commission on Dams 2000; Ennaanay 2006).

Changes in the landscape that affect annual average water yield upstream of hydropower facilities can increase or decrease hydropower production capacity. Maps of where water yield used for hydropower is produced can help avoid unintended impacts on hydropower production or help direct land use decisions that wish to maintain power production, while balancing other uses such as conservation or agriculture. Such maps can also be used to inform investments in restoration or management that downstream stakeholders, such as hydropower companies, make in hopes of improving or maintaining water yield for this important ecosystem service. In large watersheds with multiple reservoirs for hydropower production, areas upstream of power plants that sell to a higher value market will have a higher value for this service. Maps of how much value each parcel contributes to hydropower production can help managers avoid developments in the highest hydropower value areas, understand how much value will be lost or gained as a consequence of different management options, or identify which hydropower producers have the largest stake in maintaining water yield across a landscape.

Water Yield: Reservoir Hydropower Production 3.0 Beta

We are working on the next generation platform of InVEST and deploying parts of it as prototype InVEST models. Reservoir Hydropower Production has a 3.0 prototype which can be found in the Windows Start menu after the InVEST installation is complete. New features to the 3.0 version include:

- Performance improvements to the runtime of the model.
- Outputs are simplified into shapefile polygons rather than rasterized polygons. Generally the raster outputs of the ArcGIS versions of the models have a field in a shapefile that corresponds to that output.
- The ArcGIS model is run in 3 separate steps. The standalone model has a streamlined interface to run in a single step.

5.4.3 The Model

The InVEST Reservoir Hydropower model estimates the relative contributions of water from different parts of a landscape, offering insight into how changes in land use patterns affect annual surface water yield and hydropower production.

Modeling the connections between landscape changes and hydrologic processes is not simple. Sophisticated models of these connections and associated processes (such as the WEAP model) are resource and data intensive and require substantial expertise. To accommodate more contexts, for which data are readily available, InVEST maps and models the annual average water yield from a landscape used for hydropower production, rather than directly addressing the affect of LULC changes on hydropower failure as this process is closely linked to variation in water inflow on a daily to monthly timescale. Instead, InVEST calculates the relative contribution of each land parcel to annual average hydropower production and the value of this contribution in terms of energy production. The net present value of hydropower production over the life of the reservoir also can be calculated by summing discounted annual revenues.

How it Works

The model runs on a gridded map. It estimates the quantity and value of water used for hydropower production from each subwatershed in the area of interest. It has three components, which run sequentially. First, it determines the amount of water running off each pixel as the precipitation less the fraction of the water that undergoes evapotranspiration. The model does not differentiate between surface, subsurface and baseflow, but assumes that all water yield from a pixel reaches the point of interest via one of these pathways. This model then sums and averages water yield to the subwatershed level. The pixel-scale calculations allow us to represent the heterogeneity of key driving factors in water yield such as soil type, precipitation, vegetation type, etc. However, the theory we are using as the foundation of this set of models was developed at the subwatershed to watershed scale. We are only confident in the interpretation of these models at the subwatershed scale, so all outputs are summed and/or averaged to the subwatershed scale. We do continue to provide pixel-scale representations of some outputs for calibration and model-checking purposes only. **These pixel-scale maps are not to be interpreted for understanding of hydrological processes or to inform decision making of any kind.**

Second, beyond annual average runoff, it calculates the proportion of surface water that is used for hydropower production by subtracting the surface water that is consumed for other uses. Third, it estimates the energy produced by the water reaching the hydropower reservoir and the value of this energy over the reservoir's lifetime.

Water Yield Model

The water yield model is based on the Budyko curve and annual average precipitation. First, we determine annual water yield $Y(x)$ for each pixel on the landscape x as follows:

$$Y(x) = \left(1 - \frac{AET(x)}{P(x)}\right) \cdot P(x)$$

where $AET(x)$ is the annual actual evapotranspiration for pixel x and $P(x)$ is the annual precipitation on pixel x.

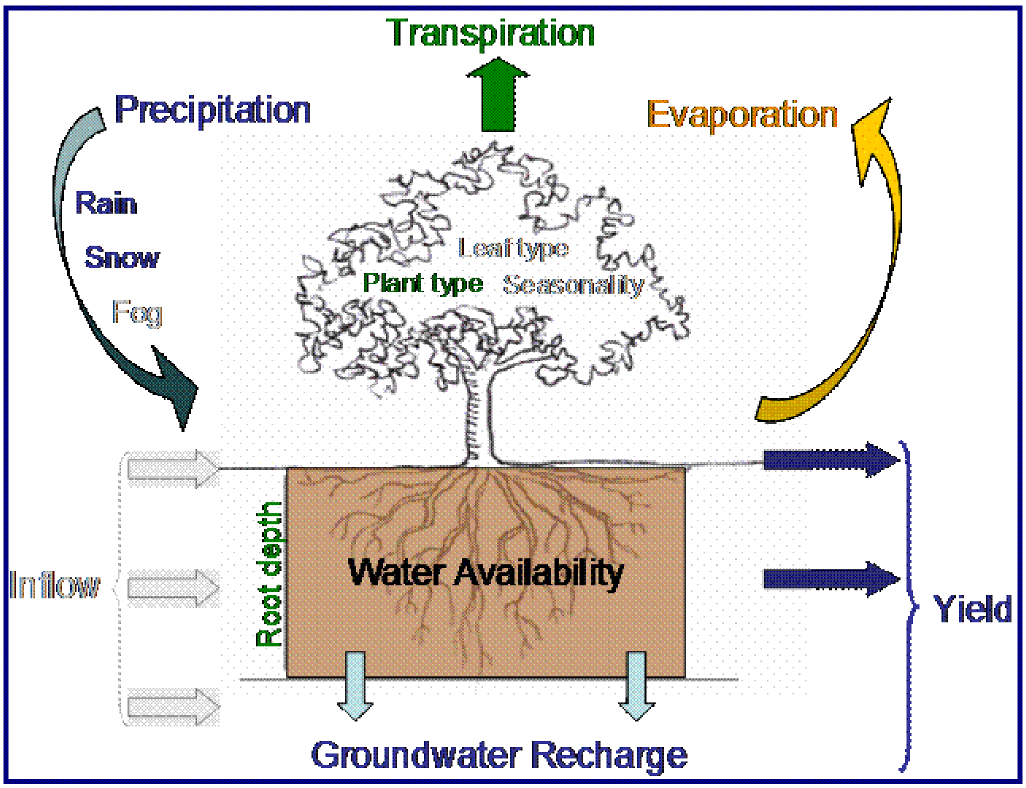


Figure 1. Conceptual diagram of the water balance model used in the hydropower production model. The water cycle is simplified, including only the parameters shown in color, and ignoring the parameters shown in gray. Yield, as calculated by this step of the model, is then adjusted for other consumptive uses and applied to hydropower energy and value estimates.

For vegetated LULC, the evapotranspiration portion of the water balance, $\frac{AET(x)}{P(x)}$, is based on an expression of the Budyko curve proposed by Fu (1981) and Zhang et al. (2004):

$$\frac{AET(x)}{P(x)} = 1 + \frac{PET(x)}{P(x)} - \left[1 + \left(\frac{PET(x)}{P(x)}\right)^{\omega}\right]^{1/\omega} \tag{5.15}$$

where $PET(x)$ is the potential evapotranspiration and $\omega(x)$ is a non-physical parameter that characterizes the natural climatic-soil properties, both detailed below.

Potential evapotranspiration $PET(x)$ is defined as:

$$PET(x) = K_c(\ell_x) \cdot ET_0(x)$$

where, $ET_0(x)$ is the reference evapotranspiration from pixel x and $K_c(\ell_x)$ is the plant (vegetation) evapotranspiration coefficient associated with the LULC ℓ_x on pixel x. $ET_0(x)$ reflects local climatic conditions, based on the evapotranspiration of a reference vegetation such as grass of alfalfa grown at that location. $K_c(\ell_x)$ is largely determined by the vegetative characteristics of the land use/land cover found on that pixel (Allen et al. 1998). K_c adjusts the ET_0 values to the crop or vegetation type in each pixel of the land use/land cover map.

$\omega(x)$ is an empirical parameter that can be expressed as linear function of $\frac{AWC*N}{P}$, where N is the number of events per year, and AWC is the volumetric plant available water content (see below for additional details). While further research is being conducted to determine the function that best describe global data, we use the expression proposed by Donohue et al. (2012) in the InVEST model, and thus define:

$$\omega(x) = Z\frac{AWC(x)}{P(x)} + 1.25$$

where:

- $AWC(x)$ is the volumetric (mm) plant available water content. The soil texture and effective rooting depth define $AWC(x)$, which establishes the amount of water that can be held and released in the soil for use by a plant. It is estimated as the product of the plant available water capacity (PAWC) and the minimum of root restricting layer depth and vegetation rooting depth:

  $$AWC(x) = Min(Rest.layer.depth, root.depth) \cdot PAWC$$

  > Root restricting layer depth is the soil depth at which root penetration is inhibited because of physical or chemical characteristics. Vegetation rooting depth is often given as the depth at which 95% of a vegetation type's root biomass occurs. PAWC is the plant available water capacity, i.e. the difference between field capacity and wilting point.

- Z is an empirical constant, sometimes referred to as "seasonality factor", which captures the local precipitation pattern and additional hydrogeological characteristics. It is positively correlated with N, the number of rain events per year. The 1.25 term is the minimum value of $\omega(x)$, which can be seen as a value for bare soil (when root depth is 0), as explained by Donohue et al. (2012). Following the literature (Yang et al., 2008; Donohue et al. 2012), values of $\omega(x)$ are capped to a value of 5.

For other LULC (open water, urban, wetland), actual evapotranspiration is directly computed from the reference evapotranspiration $ET_0(x)$ and has an upper limit defined by the precipitation:

$$AET(x) = Min(K_c(\ell_x) \cdot ET_0(x), P(x)) \tag{5.16}$$

where $ET_0(x)$ is the reference evapotranspiration, and $K_c(\ell_x)$ is the evaporation factor for each LULC. Guidance for estimating the K_c factor is provided in the "Data sources" section.

The water yield model script generates and outputs the total and average water yield at the subwatershed level.

Realized Supply Model

The Realized Supply Model calculates the water inflow to a reservoir based on water yield and water consumptive use in the watershed(s) of interest. The user inputs how much water is consumed by each land use land cover type in a table format. Examples of consumptive use include municipal or industrial withdrawals that are not returned to the stream upstream of the outlet. This model may also be used to represent inter-basin transfers out of the study watershed.

For example, in an urban area, consumptive use can be calculated as the product of population density and per capita consumptive use. These land use-based values only relate to the consumptive portion of demand; some water use is non-consumptive such as water used for industrial processes or waste water that is returned to the stream after use, upstream of the outlet. Consumptive use estimates should therefore take into account any return flows to the stream above the watershed outlet:

$$C = \frac{W - R}{n}$$

where, C = the consumptive use ($m^3/yr/pixel$), W = withdrawals (m^3/yr), R = return flows (m^3/yr), and n = number of pixels in a given land cover.

For simplicity, each pixel in the watershed is either a "contributing" pixel, which contributes to hydropower production, or a "use" pixel, which uses water for other consumptive uses. This assumption implies that land use associated with consumptive uses will not contribute any yield for downstream use. The amount of water that actually reaches the reservoir for dam d (realized supply) is defined as the difference between total water yield from the watershed and total consumptive use in the watershed:

$$V_{in} = Y - u_d$$

where V_{in} is the realized supply (volume inflow to a reservoir), u_d is the total volume of water consumed in the watershed upstream of dam d and Y is the total water yield from the watershed upstream of dam d.

Note that only anthropogenic uses are considered here, since evapotranspiration (including consumptive use of water by croplands) are accounted for by the K_c parameter in the water yield model. Users should be aware that the model assumes that all water available for evapotranspiration comes from within the watershed (as rainfall). This assumption holds true in cases where agriculture is either rain-fed, or the source of irrigation water is within the study watershed (not sourced from inter-basin transfer or a disconnected deeper aquifer). See the Limitations section for more information on applying the model in watersheds with irrigated agriculture.

If the user has observed data available on actual annual inflow rates to the reservoir for dam d, they can be compared to V_{in}.

Hydropower Production and Valuation Model

The reservoir hydropower model estimates both the amount of energy produced given the estimated realized supply of water for hydropower production and the value of that energy. A present value dollar (or other currency) estimate is given for the entire remaining lifetime of the reservoir. Net present value can be calculated if hydropower production cost data are available. The energy produced and the revenue is then redistributed over the landscape based on the proportional contribution of each subwatershed to energy production. Final output maps show how much energy production and hydropower value can be attributed to each subwatershed's water yield over the lifetime of the reservoir.

At dam d, power is calculated using the following equation:

$$p_d = \rho \cdot q_d \cdot g \cdot h_d$$

where p_d is power in watts, ρ is the water density (1000 Kg/m^3), q_d is the flow rate (m^3/s), g is the gravity constant (9.81 m/s^2), and h_d is the water height behind the dam at the turbine (m). In this model, we assume that the total annual inflow water volume is released equally and continuously over the course of each year.

The power production equation is connected to the water yield model by converting the annual inflow volume adjusted for consumption (V_{in}) to a per second rate. Since electric energy is normally measured in kilowatt-hours, the power p_d is multiplied by the number of hours in a year. All hydropower reservoirs are built to produce a maximum amount of electricity. This is called the energy production rating, and represents how much energy could be produced if the turbines are 100% efficient and all water that enters the reservoir is used for power production. In the real world, turbines have inefficiencies and water in the reservoir may be extracted for other uses like irrigation, retained in the reservoir for other uses like recreation, or released from the reservoir for non-power production uses like maintaining environmental flows downstream. To account for these inefficiencies and the flow rate and power unit adjustments, annual average energy production ε_d at dam d is calculated as follows:

$$\varepsilon_d = 0.00272 \cdot \beta \cdot \gamma_d \cdot h_d \cdot V_{in}$$

where ε_d is hydropower energy production (KWH), β is the turbine efficiency coefficient (%), γ_d is the percent of inflow water volume to the reservoir at dam d that will be used to generate energy.

To convert ε_d, the annual energy generated by dam d, into a net present value (NPV) of energy produced (point of use value) we use the following,

$$NPVH_d = (p_e\varepsilon_d - TC_d) \times \sum_{t=0}^{T-1} \frac{1}{(1+r)^t}$$

where TC_d is the total annual operating costs for dam d, p_e is the market value of electricity (per kilowatt hour) provided by hydropower plant at dam d, T_d indicates the number of years present landscape conditions are expected to persist or the expected remaining lifetime of the station at dam d (set T to the smallest value if the two time values differ), and r is the market discount rate. The form of the equation above assumes that TC_d, p_e, and ε_d, are constant over time.

Energy production over the lifetime of dam d is attributed to each subwatershed as follows:

$$\varepsilon_x = (T_d\varepsilon_d) \times (c_x/c_{tot})$$

where the first term in parentheses represents the electricity production over the lifetime of dam d. The second term represents the proportion of water volume used for hydropower production that comes from subwatershed x relative to the total water volume for the whole watershed. The value of each subwatershed for hydropower production over the lifetime of dam d is calculated similarly:

$$NPVH_x = NPVH_d \times (c_x/c_{tot})$$

Limitations and Simplifications

The model has a number of limitations. First, it is not intended for devising detailed water plans, but rather for evaluating how and where changes in a watershed may affect hydropower production for reservoir systems. It is based on annual averages, which neglect extremes and do not consider the temporal dimensions of water supply and hydropower production.

Second, the model does not consider the spatial distribution of land use land cover. The empirical model used for the water balance (based on the Budyko theory) has been tested at larger scales than the pixel dimensions used in InVEST (Hamel & Guswa, in review). Complex land use patterns or underlying geology, which may induce complex water balances, may not be well captured by the model.

Third, the model does not consider sub-annual patterns of water delivery timing. Water yield is a provisioning function, but hydropower benefits are also affected by flow regulation. The timing of peak flows and delivery of minimum operational flows throughout the year determines the rate of hydropower production and annual revenue. Changes in landscape scenarios are likely to affect the timing of flows as much as the annual water yield, and are of particular concern when considering drivers such as climate change. Modeling the temporal patterns of overland flow requires detailed data that are not appropriate for our approach. Still, this model provides a useful initial assessment of how landscape scenarios may affect the annual delivery of water to hydropower production.

Fourth, the model greatly simplifies consumptive demand. For each LULC, a single variable (γ_d) is used to represent multiple aspects of water resource allocation, which may misrepresent the complex distribution of water among uses and over time. In reality, water demand may differ greatly between parcels of the same LULC class. Much of the water demand may also come from large point source intakes, which are not represented by LULC class. The model simplifies water demand by distributing it over the landscape. For example, the water demand may be large for an urban area, and the model represents this demand by distributing it over the urban LULC class. The actual water supply intake, however, is likely further upstream in a rural location. Spatial disparity in actual and modeled demand points may cause an incorrect representation in the realized supply output grid. The distribution of consumption is also simplified in the reallocation of energy production and hydropower value since it is assumed that water consumed along flow paths is drawn equally from every pixel upstream. As a result, water scarcity, energy production patterns, and hydropower values may be incorrectly estimated.

Fifth, water transfers for irrigation, either between subbasins or between seasons, are not well captured by the model. When applying the empirical approach to cropland, one should consider the irrigation patterns, which typically fall into one of the following cases:

1. If there is no irrigation other than direct rain, one can assume that croplands respond to climate forcing in a similar way to natural vegetation (i.e. the theory behind the eco-hydrological model used in the InVEST model, linking plant available water and climate forcing, applies, cf. Donohue et al. 2012)
2. If small reservoirs store water during the wet season to irrigate crops during the dry season, the AET should equal PET during the irrigation season. However, the model predicts AET<PET due to limited water retention in undisturbed catchments (where there is no other reservoir that soil storage). This likely results in the under-estimation of evapotranspiration, and therefore the overestimation of yields. To avoid this issue, one can use the alternative equation for AET (equation 2), which sets AET directly as a function of ETo. (In that case, one should remember that AET is capped by P to avoid predicting negative water yields, which may result in an overestimation of yields).
3. If the study area contains croplands that are irrigated with water from outside the catchment (either through inter-basin transfer or pumping from a disconnected groundwater source), then AET also equals to PET during the irrigation season. Because the model assumes that evapotranspiration is sourced from rainfall, the water yield output is likely overestimated. This situation can also be represented by using the alternative equation for AET (equation 2). If one assumes that crops are being irrigated efficiently (i.e. the total volume of imported water is equal to the water deficit, or PET – P, for crop cells), then the known volume of water irrigated may be added to the modeled water yield to give a better picture of actual yield.
4. Because seasonality can play a significant role in irrigation water use, users should use caution when using the annual model in catchments with large irrigated fields. For options that are not covered above or where complex water transfers may substantially affect the water balance, users are encouraged to use alternative models that will better represent the spatial and temporal water transfers. In particular, great caution should be used when calibrating the model without good data on the different water balance components within their study area (i.e. rainfall, streamflow, irrigation rates and timing).

Finally, the model assumes that hydropower production and pricing remain constant over time. It does not account for seasonal variation in energy production or fluctuations in energy pricing, which may affect the value of hydropower. Even if sub-annual production or energy prices change, however, the relative value between parcels of land in the same drainage area should be accurate.

5.4.4 Data Needs

Here we outline the specific data used by the model. See the appendix for detailed information on data sources and pre-processing. For all raster inputs, the projection used should be defined, and the projection's linear units should be in meters.

1. **Root restricting layer depth (required).** A GIS raster dataset with an average root restricting layer depth value for each cell. Root restricting layer depth is the soil depth at which root penetration is strongly inhibited because of physical or chemical characteristics. The root restricting layer depth values should be in millimeters.

   Name: File can be named anything, but no spaces in the name and less than 13 characters if an ESRI GRID. If a TIF or IMG, the name may be longer.

   Format: Standard GIS raster file (e.g., ESRI GRID, TIF or IMG), with an average root restricting layer depth in millimeters for each cell.

Sample data set: \InVEST\Base_Data\Freshwater\depth_to_root_rest_layer

2. **Precipitation (required)**. A GIS raster dataset with a non-zero value for average annual precipitation for each cell. The precipitation values should be in millimeters.

   Name: File can be named anything, but no spaces in the name and less than 13 characters if an ESRI GRID. If a TIF or IMG, the name may be longer.

   Format: Standard GIS raster file (e.g., ESRI GRID, TIF or IMG), with precipitation values for each cell.

   Sample data set: \InVEST\Base_Data\Freshwater\precip

3. **Plant Available Water Content (required)**. A GIS raster dataset with a plant available water content value for each cell. Plant Available Water Content fraction (PAWC) is the fraction of water that can be stored in the soil profile that is available for plants' use. PAWC is a fraction from 0 to 1.

   Name: File can be named anything, but no spaces in the name and less than 13 characters if an ESRI GRID. If a TIF or IMG, the name may be longer.

   Format: Standard GIS raster file (e.g., ESRI GRID, TIF or IMG), with available water content values for each cell.

   Sample data set: \InVEST\Base_Data\Freshwater\pawc

4. **Average Annual Reference Evapotranspiration (required).** A GIS raster dataset, with an annual average evapotranspiration value for each cell. Reference evapotranspiration is the potential loss of water from soil by both evaporation from the soil and transpiration by healthy alfalfa (or grass) if sufficient water is available. The reference evapotranspiration values should be in millimeters.

   Name: File can be named anything, but no spaces in the name and less than 13 characters if an ESRI GRID. If a TIF or IMG, the name may be longer.

   Format: Standard GIS raster file (e.g., ESRI GRID, TIF or IMG), with reference evapotranspiration values for each cell.

   Sample data set: \InVEST\Base_Data\Freshwater\eto

5. **Land use/land cover (required)**. A GIS raster dataset, with an LULC code for each cell. The LULC code should be an integer.

   Name: File can be named anything, but no spaces in the name and less than 13 characters if an ESRI GRID. If a TIF or IMG, the name may be longer.

   Format: Standard GIS raster file (e.g., ESRI GRID, TIF or IMG), with an integer LULC class code for each cell (e.g., 1 for forest, 3 for grassland, etc.). These codes must match LULC codes in the Biophysical table.

   Sample data set: \InVEST\Base_Data\Freshwater\landuse_90

6. **Watersheds (required)**. A shapefile, with one polygon per watershed. This is a layer of watersheds such that each watershed contributes to a point of interest where hydropower production will be analyzed. See the Working with the DEM section for information about generating watersheds.

   Name: File can be named anything, but no spaces in the name

   Format: Shapefile (.shp)

   Rows: Each row is one watershed

   Columns: An integer field named *ws_id* is required, with a unique integer value for each watershed

   Sample data set: \InVEST\Base_Data\Freshwater\watersheds.shp

7. **subwatersheds (required)**. A shapefile, with one polygon per subwatershed within the main watersheds specified in the Watersheds shapefile. See the Working with the DEM section for information about generating subwatersheds.

   Format: Shapefile (.shp)

   Rows: Each row is one subwatershed

   Columns: An integer field named *subws_id* is required, with a unique integer value for each subwatershed

   Sample data set: \InVEST\ Base_Data\Freshwater\subwatersheds.shp

8. **Biophysical Table (required)**. A table of land use/land cover (LULC) classes, containing data on biophysical coefficients used in this tool. NOTE: these data are attributes of each LULC class rather than attributes of individual cells in the raster map.

   Sample data set: \InVEST\Base_Data\Freshwater\Water_Tables.mdb\Biophysical_Models and \InVEST\Hydropower\inputbiophysical_table.csv

   Name: Table names should only have letters, numbers and underscores, no spaces

   Format: ⋆.dbf or ⋆.mdb for ArcGIS models, the standalone model requires a .csv file

   Rows: Each row is an LULC class.

   Columns: Each column contains a different attribute of each land use/land cover class, and must be named as follows:

   1. *lucode (Land use code)*: Unique integer for each LULC class (e.g., 1 for forest, 3 for grassland, etc.), must match the LULC raster above.
   2. *LULC_desc*: Descriptive name of land use/land cover class (optional)
   3. *LULC_veg*: Contains the information on which AET equation to use (Eq. 1 or 2). Values should be 1 for vegetated land use except wetlands, and 0 for all other land uses, including wetlands, urban, water bodies, etc.
   4. *root_depth*: The maximum root depth for vegetated land use classes, given in integer millimeters. This is often given as the depth at which 95% of a vegetation type's root biomass occurs. For land uses where the generic Budyko curve is not used (i.e. where evapotranspiration is calculated from Eq. 2), rooting depth is not needed. In these cases, the rooting depth should be set to NA.
   5. K_c: The plant evapotranspiration coefficient for each LULC class, used to obtain potential evapotranspiration by using plant physiological characteristics to modify the reference evapotranspiration, which is based on alfalfa. The evapotranspiration coefficient is thus a decimal in the range of 0 to 1.5 (some crops evapotranspire more than alfalfa in some very wet tropical regions and where water is always available).

9. **Z parameter (required).** Floating point value on the order of 1 to 30 corresponding to the seasonal distribution of precipitation (see Appendix A for more information).

10. **Demand Table (required)**. A table of LULC classes, showing consumptive water use for each landuse / landcover type. Consumptive water use is that part of water used that is incorporated into products or crops, consumed by humans or livestock, or otherwise removed from the watershed water balance.

    Sample data set: \InVEST\Base_Data\Freshwater\Water_Tables.mdb\Water_Demand and \InVEST\Hydropowerinputwater_demand_table.csv

    Name: Table names should only have letters, numbers and underscores, no spaces

    Format: ⋆.dbf or ⋆.mdb for the ArcGIS version, the standalone model requires a .csv file

    Rows: Each row is a landuse / landcover class and must contain all the landcover values found in the LULC raster

Columns: Contain water demand values per LULC class and must be named as follows:

1. *lucode*: Integer value of land use/land cover class (e.g., 1 for forest, 3 for grassland, etc.), must match LULC raster, described above.
2. *demand*: The estimated average consumptive water use for each landuse / landcover type. Water use should be given in cubic meters per year for a pixel in the land use/land cover map. Note that accounting for pixel area is important since larger areas will consume more water for the same land cover type.

11. **Hydropower valuation table**. A table of hydropower stations with associated model values.

 Sample data set: \InVEST\Base_Data\Freshwater\Water_Tables.mdb\Hydropower_Valuation and \InVESTHydropowerinputhydropower_valuation_table.csv

 Name: Table names should only have letters, numbers and underscores, no spaces

 Format: *.dbf or *.mdb for the ArcGIS version, the standalone model requires a .csv file

 Rows: Each row is a hydropower station

 Columns: Each column contains an attribute of each hydropower station, and must be named as follows:

 1. *ws_id*: Unique integer value for each watershed, which must correspond to values in the Watersheds layer.
 2. *station_desc*: Name of hydropower station (optional)
 3. *efficiency*: The turbine efficiency. A number to be obtained from the hydropower plant manager (floating point values generally 0.7 to 0.9)
 4. *fraction*: The fraction of inflow water volume that is used to generate energy, to be obtained from the hydropower plant manager. Managers can release water without generating electricity to satisfy irrigation, drinking water or environmental demands. Floating point value.
 5. *height*: The head, measured as the average annual effective height of water behind each dam at the turbine intake in meters. Floating point value.
 6. *kw_price*: The price of one kilowatt-hour of power produced by the station, in dollars or other currency. Floating point value.
 7. *cost*: Annual cost of running the hydropower station (maintenance and operations costs). Floating point value.
 8. *time_span*: An integer value of either the expected lifespan of the hydropower station or the period of time of the land use scenario of interest. Used in net present value calculations.
 9. *discount*: The discount rate over the time span, used in net present value calculations. Should be represented as a percentage.

5.4.5 Running the Model

The model is available as a standalone application accessible from the Windows start menu. For Windows 7 or earlier, this can be found under *All Programs -> InVEST +VERSION+ -> Water Yield*. Windows 8 users can find the application by pressing the windows start key and typing "water" to refine the list of applications. The standalone can also be found directly in the InVEST install directory under the subdirectory *invest-3_x86/invest_hydropower_water_yield.exe*.

Viewing Output from the Model

Upon successful completion of the model, a file explorer window will open to the output workspace specified in the model run. This directory contains an *output* folder holding files generated by this model. Those files can be viewed in any GIS tool such as ArcGIS, or QGIS. These files are described below in Section *Interpreting Results*.

5.4.6 Interpreting Results

The following is a short description of each of the outputs from the Hydropower Production model. Final results are found in the *output* folder within the *workspace* specified for this model.

- **Parameter log**: Each time the model is run, a text (.txt) file will appear in the *Output* folder. The file will list the parameter values for that run and will be named according to the service, the date and time, and the suffix.
- Outputs in the *per_pixel* folder within the *output* folder can be useful for intermediate calculations but should **NOT** be interpreted at the pixel level, as model assumptions are based on processes understood at the subwatershed scale.
  - **output\per_pixel\fractp** (fraction): Estimated actual evapotranspiration fraction of precipitation per pixel (Actual Evapotranspiration / Precipitation). It is the mean fraction of precipitation that actually evapotranspires at the pixel level.
  - **output\per_pixel\aet** (mm): Estimated actual evapotranspiration per pixel.
  - **output\per_pixel\wyield** (mm): Estimated water yield per pixel.
- **output\subwatershed_results_wyield.shp** and **output\subwatershed_results_wyield.csv**: Shapefile and table containing biophysical output values per subwatershed, with the following attributes:
  - *precip_mn* (mm): Mean precipitation per pixel on the subwatershed.
  - *PET_mn* (mm): Mean potential evapotranspiration per pixel on the subwatershed.
  - *AET_mn* (mm): Mean actual evapotranspiration per pixel on the subwatershed.
  - *wyield_mn* (mm): Mean water yield per pixel on the subwatershed.
  - *num_pixels*: Number of pixels per subwatershed.
  - *wyield_vol* (m^3): Volume of water yield in the subwatershed.
  - *wyield_ha* (m^3): Volume of water yield in the subwatershed per hectare.
- **output\watershed_results_wyield.shp** and **output\watershed_results_wyield.csv**: Shapefile and table containing output values per watershed:

  When the water yield model is run, the following biophysical outputs result:
  - *precip_mn* (mm): Mean precipitation per pixel on the watershed.
  - *PET_mn* (mm): Mean potential evapotranspiration per pixel on the watershed.
  - *AET_mn* (mm): Mean actual evapotranspiration per pixel on the watershed.
  - *wyield_mn* (mm): Mean water yield per pixel on the watershed.
  - *num_pixels*: Number of pixels per watershed.
  - *wyield_vol* (m^3): Volume of water yield in the watershed.
  - *wyield_ha* (m^3): Volume of water yield in the watershed per hectare.

  If the water scarcity model is run, the following attributes will also be included:
  - **consum_vol** (m^3): Total water consumption for each watershed.

- **consum_mn** (m^3/ha): Mean water consumptive volume per hectare per watershed.
- **rsupply_vl** (m^3): Total realized water supply (water yield – consumption) volume for each watershed.
- **rsupply_mn** (m^3/ha): Mean realized water supply (water yield – consumption) volume per hectare per watershed.

If the hydropower production and valuation model is run, the following attributes will also be included:

- **hp_energy** (kw/timespan): THIS IS THE AMOUNT OF THIS ECOSYSTEM SERVICE IN ENERGY PRODUCTION TERMS. This grid shows the amount of energy produced by the hydropower station over the specified timespan that can be attributed to each watershed based on its water yield contribution.
- **hp_val** (currency/timespan): THIS IS THE VALUE OF THIS ECOSYSTEM SERVICE IN ECONOMIC TERMS. This grid shows the value of the landscape per watershed according to its ability to yield water for hydropower production over the specified timespan.

The application of these results depends entirely on the objective of the modeling effort. Users may be interested in all of these results or a select one or two. If costing information is not available or of interest, the user may choose to simply run the water yield model and compare biophysical results.

The first several model results provide insight into how water is distributed through the landscape. *aet_mn* describes the actual evapotranspiration depth of the hydrologic cycle, showing how much water (precipitation) is lost annually per pixel to evapotranspiration across the subwatershed or subwatershed.

The *wyield_vol* field contains the estimated annual average water volume that is 'yielded' from each subwatershed of the watershed of interest. This value can be used to determine which subwatersheds are most important to total annual water yield – although at this step the user still will not know how much of that water is benefiting downstream users of any type. The consumptive use (*consum_vol*) field then shows how much water is used for consumptive activities (such as drinking, bottling, etc.) each year across the landscape per watershed. The realized supply (*rsupply_vl*) field contains the difference between cumulative water yield and cumulative consumptive use. This value demonstrates where the water supply for hydropower production is abundant and where it is most scarce. The user needs to remember that the consumptive use value may not truly represent where water is taken, only where it is demanded. This may cause some misrepresentation of the scarcity in certain locations, but this value offers a general sense of the water balance and whether there is a lack of or abundance of water in the watershed of interest.

The *hp_energy* and *hp_val* values are the most relevant model outputs for prioritizing the landscape for investments that wish to maintain water yield for hydropower production. The *hp_val* field contains the most information for this purpose as it represents the revenue attributable to each watershed over the expected lifetime of the hydropower station, or the number of years that the user has chosen to model. This value accounts for the fact that different hydropower stations within a large river basin may have different customers who pay different rates for energy production. If this is the case, this grid will show which watersheds contribute the highest value water for energy production. If energy values do not vary much across the landscape, the *hp_energy* outputs can be just as useful in planning and prioritization. Comparing any of these values between landuse scenarios allows the user to understand how the role of the landscape may change under different management plans.

5.4.7 Appendix A: Data Sources

This is a rough compilation of data sources and suggestions about finding, compiling, and formatting data. This section should be used for ideas and suggestions only. We will continue to update this section as we learn about new data sources and methods. A good resource for any type of free geographic datasets is: http://freegisdata.rtwilson.com/

1. **Average annual precipitation**

   Average Annual Precipitation may be interpolated from existing rain gages, and global data sets from remote sensing models to account for remote areas. Precipitation as snow is included. If field data are not available, you can use coarse data from the freely available global data set developed by the Climatic Research Unit (http://www.cru.uea.ac.uk).

Within the United States, the PRISM group at Oregon State University provides free precipitation data at a 30-arcsecond resolution. See their website at http://www.prism.oregonstate.edu/ and navigate to '800 m Normals' to download data.

2. **Average annual reference evapotranspiration** (ET_0)

   Reference evapotranspiration, ET_0, is the energy (expressed as a depth of water, e.g. mm) supplied by the sun (and occasionally wind) to vaporize water. Some global products are available on the internet, such as FAO Penman-Monteith method with limited climatic data, as described in FAO Irrigation and Drainage Paper 56 using data from the Climatic Research Unit. Reference evapotranspiration varies with elevation, latitude, humidity, and slope aspect. There are countless methodologies, which range in data requirements and precision.

   If the use of this grid is not possible, develop monthly average grids of precipitation, and maximum and minimum temperatures (http://www.cru.uea.ac.uk), which need to incorporate the effects of elevation when interpolating from observation stations. Data to develop these monthly precipitation and temperatures grids follow the same process in the development of the 'Average Annual Precipitation' grid, with the added monthly disaggregated grids.

   A simple way to determine reference Evapotranspiration is the 'modified Hargreaves' equation (Droogers and Allen, 2002), which generates superior results than the Pennman-Montieth when information is uncertain.

   $$:math: `ET_0` = 0.0013 \times 0.408 \times RA \times (T_{av} + 17) \times (TD - 0.0123P)^{0.76}$$

   The 'modified Hargreaves' uses the average of the mean daily maximum and mean daily minimum temperatures (Tavg in oC), the difference between mean daily maximum and mean daily minimums (TD), RA is extraterrestrial radiation (RA in $\mathrm{MJm^{-2}d^{-1}}$ and precipitation (P in mm per month), all of which can be relatively easily obtained. Temperature and precipitation data are often available from regional charts or direct measurement. Radiation data, on the other hand, is far more expensive to measure directly but can be reliably estimated from online tools, tables or equations.

   The reference evapotranspiration could be also calculated monthly and annually using the Hamon equation (Hamon 1961, Wolock and McCabe 1999):

   $$PED_{Hamon} = 13.97dD^2W_t$$

   where *d* is the number of days in a month, *D* is the mean monthly hours of daylight calculated for each year (in units of 12 hours), and Wt is a saturated water vapor density term calculated by:

   $$W_t = \frac{4.95e^{0.062T}}{100}$$

   where T is the monthly mean temperature in degrees Celsius. Reference evapotranspiration is set to zero when mean monthly temperature is below zero. Then for each year during the time periods analyzed, the monthly calculated PET values at each grid cell are summed to calculate a map of the annual PET for each year.

   A final method to assess ETo, when pan evaporation data are available, is to use the following equation (). ETo = pan ET *0.7 (Allen et al., 1998)

3. **Root restricting layer depth**

   Root restricting layer depth is the soil depth at which root penetration is strongly inhibited because of physical or chemical characteristics. Root restricting layer depth may be obtained from some soil maps. If root restricting layer depth or rootable depth by soil type is not available, soil depth can be used as a proxy. The FAO provides global soil data in their Harmonized World

Soil Database: http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/ Soil data for many parts of the world are also available from the Soil and Terrain Database (SOTER) Programme: http://www.isric.org/projects/soil-and-terrain-database-soter-programme.

In the United States free soil data is available from the U.S. Department of Agriculture's NRCS in the form of two datasets: SSURGO http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs142p2_053627 and STATSGO http://water.usgs.gov/GIS/metadata/usgswrd/XML/ussoils.xml . Where available SSURGO data should be used, as it is much more detailed than STATSGO. Where gaps occur in the SSURGO data, STATSGO can be used to fill in the blanks. If several soil horizons are detailed, the root restricting layer depth is the sum of the depths of non-restrictive soil horizons. The Soil Data Viewer (http://www.nrcs.usda.gov/wps/portal/nrcs/detailfull/soils/home/?cid=nrcs142p2_053620) can be used for soil data processing and should be used whenever possible.

Ultimately, a grid layer must be produced.

4. **Plant available water content (PAWC)**

   Plant available water content is a fraction obtained from some standard soil maps. It is defined as the difference between the fraction of volumetric field capacity and permanent wilting point. Often plant available water content is available as a volumetric value (mm). To obtain the fraction divide by soil depth. Soil characteristic layers are estimated by performing a weighted average from all horizons within a soil component. If PAWC is not available, raster grids obtained from polygon shape files of weight average soil texture (%clay, %sand, %silt) and soil porosity will be needed. See 'Root Restricting Layer Depth' above for a description of where to find and how to process soil data. http://hydrolab.arsusda.gov/SPAW/Index.htm has software to help you estimate your PAWC when you have soil texture data.

5. **Land use/land cover**

   A key component for all Tier 1 water models is a spatially continuous landuse / land class raster grid. That is, within a watershed, all landuse / land class categories should be defined. Gaps in data that break up the drainage continuity of the watershed will create errors. Unknown data gaps should be approximated. Global land use data is available from the University of Maryland's Global Land Cover Facility: http://glcf.umd.edu/data/landcover/. This data is available in 1 degree, 8km and 1km resolutions. Multi-year global landcover data is provided in several different classifications in the MODIS Land Cover from NASA: https://lpdaac.usgs.gov/products/modis_products_table/mcd12q1. The European Space Agency provides landcover maps for 2005 and 2009 at http://due.esrin.esa.int/globcover/. Data for the U.S. for 1992 and 2001 is provided by the EPA in their National Land Cover Data product: http://www.epa.gov/mrlc/.

   The simplest categorization of LULCs on the landscape involves delineation by land cover only (e.g., cropland, temperate conifer forest, and prairie). Several global and regional land cover classifications are available (e.g., Anderson et al. 1976), and often detailed land cover classification has been done for the landscape of interest.

   A slightly more sophisticated LULC classification could involve breaking relevant LULC types into more meaningful categories. For example, agricultural land classes could be broken up into different crop types or forest could be broken up into specific species.

   The categorization of land use types depends on the model and how much data is available for each of the land types. The user should only break up a land use type if it will provide more accuracy in modeling. For instance, for the water quality model the user should only break up 'crops' into different crop types if they have information on the difference in nutrient loading between crops. Along the same lines, the user should only break the forest land type into specific species for the water supply model if information is available on the root depth and evapotranspiration coefficients for the different species.

   Sample Land Use/Land Cover Table

ID	Land Use/Land Cover
1	Evergreen Needleleaf Forest
2	Evergreen Broadleaf Forest
3	Deciduous Needleleaf Forest
4	Deciduous Broadleaf Forest
5	Mixed Cover
6	Woodland
7	Wooded Grassland
8	Closed Shrubland
9	Open Shrubland
10	Grassland
11	Cropland (row Crops)
12	Bare Ground
13	Urban and Built-Up
14	Wetland
15	Mixed evergreen
16	Mixed Forest
17	Orchards/Vineyards
18	Pasture

6. **Root depth**

   A valuable review of plant rooting depths was done by Schenk and Jackson (2002). Root depth values should be based on depth at which 90% of root biomass occurs, not the maximum depth of the longest tap root. Other rooting depth values for crops and some tree plantations can be found in the FAO 56 guidelines by Allen et al. (1998).

   The model determines the minimum of root restricting layer depth and rooting depth for an accessible soil profile for water storage. Values must be integer, converted to mm. For non-vegetated LULCs (e.g. urban), for which Equation 2 above is used, the model will not use the root depth value so any value can be inserted in the table.

7. **Evapotranspiration coefficient table Kc**

   Evapotranspiration coefficient (K_c) values for crops are readily available from irrigation and horticulture handbooks. FAO has an online resource for this: http://www.fao.org/docrep/X0490E/x0490e0b.htm. The FAO tables list coefficients by crop growth stage (K_c ini, K_c mid, K_c end), which need to be converted to an annual average K_c because this is an annual water yield model. This requires knowledge about the phenology of the vegetation in the study region (average green-up, die-down dates) and crop growth stages (when annual crops are planted and harvested). Annual average K_c can be estimated as a function of vegetation characteristics and average monthly reference evapotranspiration using the following equation:

$$K_c = \frac{\sum_{m=1}^{12} K_{cm} \times ET_{o_m}}{\sum_{m=1}^{12} ET_{o_m}}$$

   where K_{cm} is an average crop coefficient of month m (1-12) and ET_{o_m} is the corresponding reference evapotranspiration. These values can also be calculated using the following spreadsheet: http://ncp-dev.stanford.edu/~dataportal/invest-data/Kc_calculator.xlsx. Values for K_c should be decimals between 0-1.5.

   Values for other vegetation can be estimated using Leaf Area Index (LAI) relationships. LAI characterizes the area of green leaf per unit area of ground surface and can be obtained by satellite imagery products derived from NDVI analysis. A typical LAI - K_c relationship is as follows (Allen et al., 1998, Chapter 6: http://www.fao.org/docrep/x0490e/x0490e0b.htm):

$$K_c = \begin{cases} \frac{LAI}{3} \text{ when } LAI \leq 3 \\ 1 \end{cases}$$

K_c estimates for non-vegetated LULC are based on (Allen et al., 1998). Note that these values are only approximate, but unless the LULC represents a significant portion of the watershed, the impact of the approximation on model results should be minimal.

- Kc for <2m open water can be approximated by Kc=1;
- Kc for >5m open water is in the range of 0.7 to 1.1;
- Kc for wetlands can be assumed in the range of 1 to 1.2;
- Kc for bare soil ranges from 0.3 to 0.7 depending on climate (in particular rainfall frequency). It can be estimated at Kc=0.5 (see Allen 1998, Chapter 11). Additional information for determining Kc for bare soil can be found in (Allen et al., 2005).
- Kc for built areas can be set to f*0.1 +(1-f)*0.6 where f is the fraction of impervious cover in the area. Here, evapotranspiration from pervious areas in built environments is assumed to be approximately 60% of reference evapotranspiration (i.e. the average between lawn grass and bare soil). In addition, evaporation from impervious surface is assumed at 10% of PET. Should local data be available, the user may compute an annual average estimate of Kc, using the method described for crop factors.

No zero values are allowed.

Sample Evapotranspiration coefficient Kc Table.

ID	Vegetation Type	Kc
1	Evergreen Needleleaf Forest	1
2	Evergreen Broadleaf Forest	1
3	Deciduous Needleleaf Forest	1
4	Deciduous Broadleaf Forest	1
5	Mixed Cover	1
6	Woodland	1
7	Wooded Grassland	1
8	Closed Shrubland	0.398
9	Open Shrubland	0.398
10	Grassland	0.65
11	Cropland (Row Crops)	0.65
12	Bare Ground	0.5
13	Urban and Built-Up	0.3
14	Wetland	1.2
15	Mixed Evergreen	1
16	Mixed Forest	1
17	Orchards/Vineyards	0.7
18	Pasture	0.85
19	Sclerophyllous Forests	1

8. **Digital elevation model (DEM)**

   DEM data is available for any area of the world, although at varying resolutions. Free raw global DEM data is available on the internet from NASA - http://asterweb.jpl.nasa.gov/gdem.asp, and USGS - http://eros.usgs.gov/elevation-products and http://hydrosheds.cr.usgs.gov/. Or a final product may be purchased relatively inexpensively at sites such as MapMart (www.mapmart.com). The DEM used in the model must be hydrologically correct meaning that sinks are filled and there are no holes. See the Working with the DEM section of this manual for more information.

1. **Consumptive water use**

   The consumptive water use for each land use / land class type is the water that is removed from the water balance. It should be estimated based on knowledge of local water transfers (e.g. extraction from groundwater or surface water for urban water supply) in consultation with local professionals in these fields. The value used in the table is an average for each land use type. For agricultural areas, water used by cattle or agricultural processing that is not returned to the watershed must be considered. In urban areas, water use may be calculated based on an estimated water use per person and multiplied by the approximate population area per raster cell. Industrial water use or water exports to other watersheds must also be considered where applicable. For all of these calculations, it is assumed that the agricultural water demand, people, etc. are spread evenly across each land use class.

10. **Hydropower Watersheds and subwatersheds**

    See the Working with the DEM section of this manual for information on generating watersheds and subwatersheds.

    The resulting delineation should be checked to ensure that the watersheds accurately represent reality. This reality check may involve talking to a local hydrologist, checking the drainage area for a nearby USGS gage, or doing a back of the envelope calculation for the annual rainfall multiplied by the watershed area and comparing it to the average annual volume of flow into the hydropower station.

    If you do not have a starting point for subwatersheds, the global dataset from Hydro1k may be applicable: http://lta.cr.usgs.gov/HYDRO1K.

11. **Hydropower Station Information**

    Detailed information about each hydropower station may only be available from the owner or managing entity of the stations. Some information may be available through public sources, and may be accessible online. In particular, if the hydropower plant is located in the United States information may be found on the internet. The first place to check is the National Inventory of Dams (http://geo.usace.army.mil/pgis/f?p=397:1:0). If a hydropower dam is owned by the Bureau of Reclamation, they should have information on the reservoir on their Dataweb (http://www.usbr.gov/projects/). Similar information may be found online at other websites for reservoirs owned or operated by other government agencies or energy companies.

    Global collections of dam locations and information include the Global Reservoir and Dam (GRanD) Database (http://www.gwsp.org/products/grand-database.html) and the World Water Development Report II dam database (http://wwdrii.sr.unh.edu/download.html.)

    - *Calibration*: For calibration, data are needed on how much water actually reaches the (sub)watershed outlets, which can be a hydropower station, on an average annual basis. Data should be available from the managing entity of the hydropower plant. In absence of information available directly from the hydropower operators, data may be available for a stream gage just upstream of the hydropower station. Gages in the U.S. may be managed by the USGS, the state fish and wildlife agency, the state department of ecology or by a local university.
    - *Time_period*: The design life span of each hydropower station can be obtained from the station owner or operator. Alternative sources may be available online as described above.

    This value may instead represent the time period of a scenario of interest, which should be equal to or smaller than the life span of the station.

    - *Discount_rate*: this rate is defined as how much value the currency loses per year.

12. **Z parameter**

Z is an empirical constant that captures the local precipitation pattern and hydrogeological characteristics, with typical values ranging from 1 to 30. Several studies have determined ω empirically (e.g. Xu et al. 2013, Fig. 3; Liang and Liu 2014; Donohue et al. 2012) and can be used to estimate Z. The relationship between ω and Z is:

$$Z = \frac{(\omega - 1.25)P}{AWC}$$

where P and AWC should be average values of Precipitation and Available Water Capacity, respectively, in the study area. AWC is the volumetric (mm) plant available water content. The soil texture and effective rooting depth define

AWC, which establishes the amount of water that can be held and released in the soil for use by a plant. It is estimated as the product of the plant available water capacity (PAWC) and the minimum of root restricting layer depth and vegetation rooting depth:

$$AWC = Min(Rest.layer.depth, root.depth) \cdot PAWC$$

Root restricting layer depth is the soil depth at which root penetration is inhibited because of physical or chemical characteristics. Vegetation rooting depth is often given as the depth at which 95% of a vegetation type's root biomass occurs. PAWC is the plant available water capacity, i.e. the difference between field capacity and wilting point.

Alternatively, following a study by Donohue et al. (2012) encompassing a range of climatic conditions in Australia, Z could be estimated as 0.2*N, where N is the number of rain events per year. The definition of a rain event is the one used by the authors of the study, characterized by a minimum period of 6 hours between two storms. Calibration of the Z coefficient may also be used by comparing modeled and observed data. Note that the Budyko curve theory suggests that the sensitivity of the model to Z is lower when Z values are high, or in areas with a very low or very high aridity index ($\frac{ET_0}{P}$; see Fig. 5 in Zhang et al. 2004).

5.4.8 Appendix B: Calibration of Water Yield Model

The water yield model is based on a simple water balance where it is assumed that all water in excess of evaporative loss arrives at the outlet of the watershed. The model is an annual average time step simulation tool applied at the pixel level but reported at the subwatershed level. If possible, calibration of the model should be performed using long term average streamflow (as a rule of thumb, a 10-year period can be used to capture some climate variability). Gauge data is often provided in flow units (i.e m^3/s). Since the model calculates water volume, the observed flow data should be converted into units of m^3/year. Climate data (total precipitation and potential evapotranspiration) should also match the date of the land use map. The other inputs, root restricting layer depth and plant available water content are less susceptible to temporal variability so any available data for these parameters may be used.

As with all models, model uncertainty is inherent and must be considered when analyzing results for decision making. Before the user starts the calibration process, we highly recommend conducting sensitivity analyses. The sensitivity analyses will define the parameters that influence model outputs the most (see for example Hamel and Guswa, in review; Sanchez-Canales et al., 2012). The calibration can then focus on highly sensitive parameters.

5.4.9 References

Allen, R.G., Pereira, L.S., Raes, D. and Smith, M., 1998. "Crop evapotranspiration. Guidelines for computing crop water requirements." FAO Irrigation and Drainage Paper 56. Food and Agriculture Organization of the United Nations, Rome, Italy. Available at: http://www.fao.org/docrep/x0490e/x0490e00.htm

Allen, R., Pruitt, W., Raes, D., Smith, M. and Pereira, L., 2005. "Estimating Evaporation from Bare Soil and the Crop Coefficient for the Initial Period Using Common Soils Information." Journal of Irrigation and Drainage Engineering, 131(1): 14-23.

Donohue, R. J., M. L. Roderick, and T. R. McVicar (2012), Roots, storms and soil pores: Incorporating key ecohydrological processes into Budyko's hydrological model, Journal of Hydrology, 436-437, 35-50

Droogers, P. & Allen, R.G. 2002. "Estimating reference evapotranspiration under inaccurate data conditions." Irrigation and Drainage Systems, vol. 16, Issue 1, February 2002, pp. 33–45

Ennaanay, Driss. 2006. Impacts of Land Use Changes on the Hydrologic Regime in the Minnesota River Basin. Ph.D. thesis, graduate School, University of Minnesota.

Fu, B. P. (1981), On the calculation of the evaporation from land surface (in Chinese), Sci. Atmos. Sin., 5, 23– 31.

Hamel, P., & Guswa, A. (in review). Uncertainty analysis of a spatially-explicit annual water-balance model: case study of the Cape Fear catchment, NC. Hydrology and Earth System Sciences.

Liang, L., & Liu, Q. (2014). Streamflow sensitivity analysis to climate change for a large water-limited basin. Hydrological Processes, 28(4), 1767–1774. doi:10.1002/hyp.9720

Sánchez-Canales, M., López Benito, A., Passuello, A., Terrado, M., Ziv, G., Acuña, V., Elorza, F. J. (2012). Sensitivity analysis of ecosystem service valuation in a Mediterranean watershed. Science of the Total Environment, 440, 140–53. doi:10.1016/j.scitotenv.2012.07.071

Schenk, H. J., & Jackson, R. B. (2002). Rooting depths, lateral root spreads and below-ground/above-ground allometries of plants in water-limited ecosystems. Journal of Ecology, 90(3), 480–494. doi:10.1046/j.1365-2745.2002.00682.x

World Commission on Dams (2000). Dams and development: A new framework for decision- making. The Report of the World Commission on Dams. Earthscan Publications LTD, London.

Xu, X., Liu, W., Scanlon, B. R., Zhang, L., & Pan, M. (2013). Local and global factors controlling water-energy balances within the Budyko framework. Geophysical Research Letters, 40(23), 6123–6129. doi:10.1002/2013GL058324

Yang, H., Yang, D., Lei, Z., & Sun, F. (2008). New analytical derivation of the mean annual water-energy balance equation. Water Resources Research, 44(3), n/a–n/a. doi:10.1029/2007WR006135

Zhang, L., Hickel, K., Dawes, W. R., Chiew, F. H. S., Western, A. W., Briggs, P. R. (2004) A rational function approach for estimating mean annual evapotranspiration. Water Resources Research. Vol. 40 (2)

5.5 Nutrient Delivery Ratio model

5.5.1 Summary

The objective of the InVEST nutrient delivery model is to map nutrient sources from watersheds and their transport to the stream. This spatial information can be used to assess the service of nutrient retention by natural vegetation. The retention service is of particular interest for surface water quality issues and can be valued in economic or social terms (e.g. avoided treatment costs, improved water security through access to clean drinking water).

The main differences between the NDR model and the InVEST v3.1 Nutrient retention model are: - The routing of nutrient from a pixel to the stream was modified to reduce the sensitivity to grid resolution and facilitate the selection of LULC-specific retention coefficient; - It is now possible to calibrate the model based on one (non-physical) parameter; note that calibration preserves the spatial distribution of nutrient sinks and sources, increasing confidence in spatially explicit outputs; - The flexible model structure allows advanced users to represent more complex processes such as direct nutrient discharges (for example, tile drainage), or instream retention (work in progress)

5.5.2 Introduction

Land use change, and in particular the conversion to agricultural lands, dramatically modifies the natural nutrient cycle. Anthropogenic nutrient sources include point sources, e.g. industrial effluent or water treatment plant discharges, and non-point sources, e.g. fertilizer used in agriculture and residential areas. When it rains or snows, water flows over the landscape carrying pollutants from these surfaces into streams, rivers, lakes, and the ocean. This has consequences for people, directly affecting their health or well-being (Keeler et al., 2012), and for aquatic ecosystems that have a limited capacity to adapt to these nutrient loads.

One way to reduce non-point source pollution is to reduce the amount of anthropogenic inputs (i.e. fertilizer management). When this option fails, ecosystems can provide a purification service by retaining or degrading pollutants before they enter the stream. For instance, vegetation can remove pollutants by storing them in tissue or releasing them back to the environment in another form. Soils can also store and trap some soluble pollutants. Wetlands can slow flow long enough for pollutants to be taken up by vegetation. Riparian vegetation is particularly important in this regard, often serving as a last barrier before pollutants enter a stream.

Land-use planners from government agencies to environmental groups need information regarding the contribution of ecosystems to mitigating water pollution. Specifically, they require spatial information on nutrient export and areas with highest filtration. The nutrient delivery and retention model provides this information for non-point source pollutants. The model was designed for nutrients (nitrogen and phosphorous), but its structure can be used for other contaminants (persistent organics, pathogens etc.) if data are available on the loading rates and filtration rates of the pollutant of interest.

5.5.3 The Model

Overview

The model uses a mass balance approach, describing the movement of mass of nutrient through space. Unlike more sophisticated nutrient models, the model does not represent the details of the nutrient cycle but rather represents the long-term, steady-state flow of nutrients through empirical relationships. Sources of nutrient across the landscape, also called nutrient loads, are determined based on the LULC map and associated loading rates. Nutrient loads can then be divided into sediment-bound and dissolved parts, which will be transported through surface and subsurface flow, respectively. Note that this step is optional; the user can choose to model surface flow only. In a second step, delivery factors are computed for each pixel based on the properties of pixels belonging to the same flow path (in particular their slope and retention efficiency of the land use). At the watershed/subwatershed outlet, the nutrient export is computed as the sum of the pixel-level contributions.

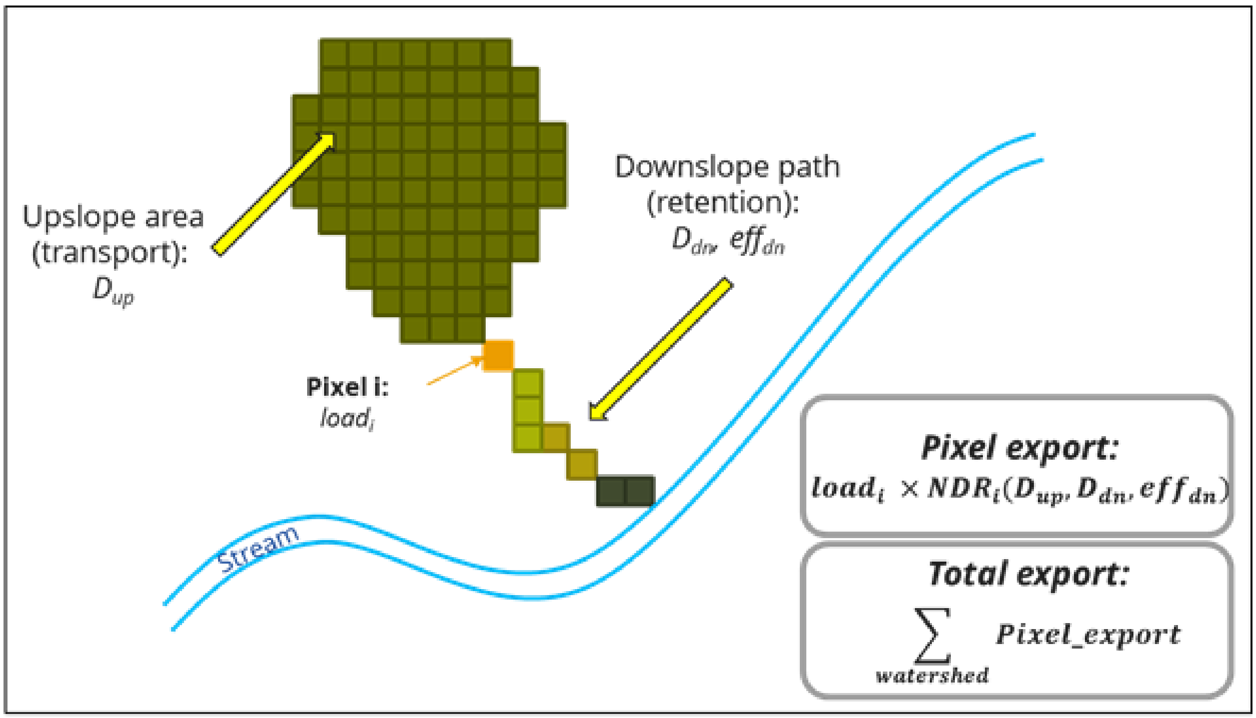


Fig. 5.1: Conceptual representation of the NDR model. Each pixel i is characterized by its nutrient load, loadi, and its nutrient delivery ratio (NDR), which is a function of the upslope area, and downslope flow path (in particular the retention efficiencies of LULC types on the downslope flow path). Pixel-level export is computed based on these two factors, and the sediment export at the watershed level is the sum of pixel-level nutrient exports.

Nutrient Loads

Loads are the sources of nutrients associated to each pixel of the landscape. Consistent with the export coefficient literature (California Regional Water Quality Control Board Central Coast Region, 2013; Reckhow et al., 1980), load values for each LULC are derived from empirical measures of nutrient export (e.g. nutrient export running off urban areas, crops, etc.). If information is available on the amount of nutrient applied (e.g. fertilizer, livestock waste, atmospheric deposition), it is possible to use it by estimating the on-pixel nutrient use (and apply this correction factor to obtain the load parameters).

Next, each pixel's load is modified to account for the local runoff potential. The LULC-based loads defined above are averages for the region, but each pixel's contribution will depend on the amount of runoff transporting nutrients (Endreny and Wood, 2003; Heathwaite et al., 2005). As a simple approximation, the loads can be modified as follows:

$$modified.load_(x,i) = load_(x,i) RPI_i \tag{5.17}$$

where RPI_i is the runoff potential index on pixel i. It is defined as: $RPI_i = RP_i/RP_av$, where RP_i is the runoff proxy for runoff on pixel i, and RP_av is the average RP over the raster. This approach is similar to that developed by Endreny and Wood (2003). In practice, the raster RP is defined either as a quickflow index (e.g. from the InVEST seasonal water yield model) or as precipitation.

For each pixel, modified loads can be divided into sediment-bound and dissolved nutrient portions. Conceptually, the former represents nutrients that are transported by surface or shallow subsurface runoff, while the latter represent nutrients transported by groundwater. The ratio between these two types of nutrient sources is given by the parameter proportion_subsurface_x (where x=n or x=p, for nitrogen or phosphorus, respectively), which quantifies the ratio of dissolved nutrients over the total amount of nutrients. For a pixel i:

$$load_{surf,i} = (1 - proportion\_subsurface_i) \cdot modified.load\_x_i \tag{5.18}$$

$$load_{subsurf,i} = proportion\_subsurface_i \cdot modified.load\_x_i \tag{5.19}$$

In case no information is available on the partitioning between the two types, the recommended default value of *load_subsurface_x* is 0, meaning that all nutrients are reaching the stream via surface flow. (Note that surface flow can, conceptually, include or shallow subsurface flow). However, users should explore the model's sensitivity to this value to characterize the uncertainty introduced by this assumption.

Nutrient Delivery

Nutrient delivery is based on the concept of nutrient delivery ratio (NDR), an approach inspired by the peer-reviewed concept of sediment delivery ratio (see InVEST sediment model user's guide and Vigiak et al., 2012). The concept is similar to the risk-based index approaches that are popular for nutrient modeling (Drewry et al., 2011), although it provides quantitative values of sediment export (e.g. the proportion of the nutrient load that will reach the stream). Two delivery ratios are computed, one for nutrient transported by surface flow, the other for subsurface flow.

Surface NDR

The surface NDR is the product of a delivery factor, representing the ability of downstream pixels to transport nutrient without retention, and a topographic index, representing the position on the landscape. For a pixel i:

$$NDR_i = NDR_{0,i}\left(1+\exp\left(\frac{IC_i - IC_0}{k}\right)\right)^{-1} \tag{5.20}$$

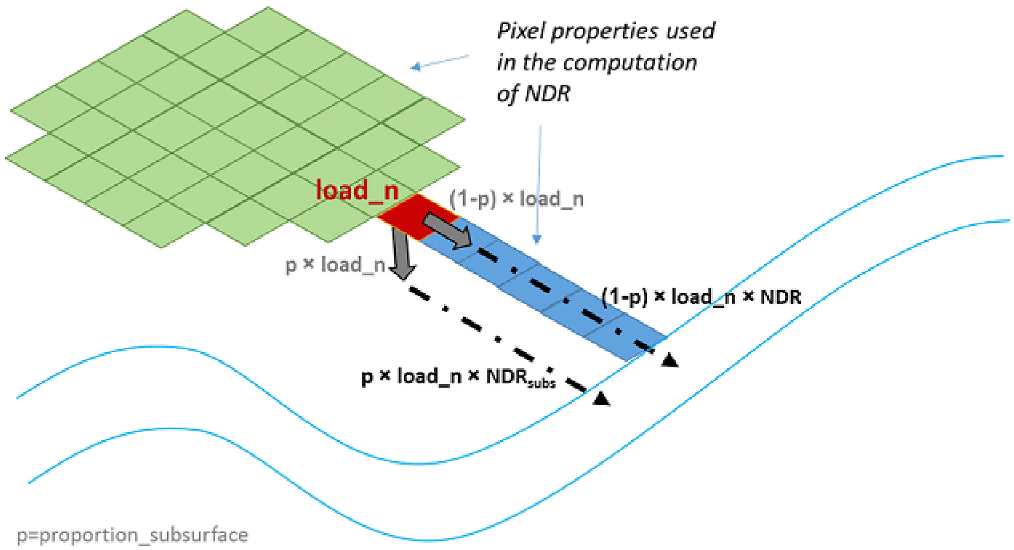


Fig. 5.2: Conceptual representation of nutrient delivery in the model. If the user chooses to represent subsurface flow, the load on each pixel, load_n, is divided into two parts, and the total nutrient export is the sum of the surface and subsurface contributions.

where IC_0 and k are calibration parameters, IC_i is a topographic index, and $NDR_{0,i}$ is the proportion of nutrient that is not retained by downstream pixels (irrespective of the position of the pixel on the landscape). Below we provide details on the computation of each factor.

$NDR_{0,i}$ is based on the maximum retention efficiency of the land between a pixel and the stream (downslope path, in Figure 1):

$$NDR_{0,i} = 1 - eff'_i \tag{5.21}$$

Moving along a flow path, the algorithm computes the additional retention provided by each pixel, taking into account the total distance traveled across each LULC type. Each additional pixel from the same LULC type will contribute a smaller value to the total retention, until the maximum retention efficiency for the given LULC is reached (Figure 2). The total retention is capped by the maximum retention value that LULC types along the flow path can provide, eff_{LULC_i}.

In mathematical terms:

$$eff'_i = \begin{cases} eff_{LULC_i} \cdot (1 - s_i) & \text{if } down_i \text{ is a stream pixel} \\ eff'_{down_i} \cdot s_i + eff_{LULC_i} \cdot (1 - s_i) & \text{if } eff_{LULC_i} > eff'_{down_i} \\ eff'_{down_i} & otherwise \end{cases}$$

Where:

- eff'_{down_i} is the effective downstream retention on the pixel directly downstream from i,
- eff_{LULC_i} is the maximum retention efficiency that LULC type i can reach, and
- s_i is the step factor defined as:

$$s_i = \exp\left(\frac{-5\ell_{i_{down}}}{\ell_{LULC_i}}\right) \tag{5.22}$$

With:

- $\ell_{i_{down}}$ is the length of the flow path from pixel i to its downstream neighbor

- ℓ_{LULC_i} is the LULC retention length of the landcover type on pixel i

Notes:

Since eff'_i is dependent on the pixels downstream, calculation proceeds recursively starting at pixels that flow directly into streams before upstream pixels can be calculated.

In equation [6], the factor 5 is based on the assumption that maximum efficiency is reached when 99% of its value is reached (assumption due to the exponential form of the efficiency function, which implies that the maximum value cannot be reached with a finite flow path length).

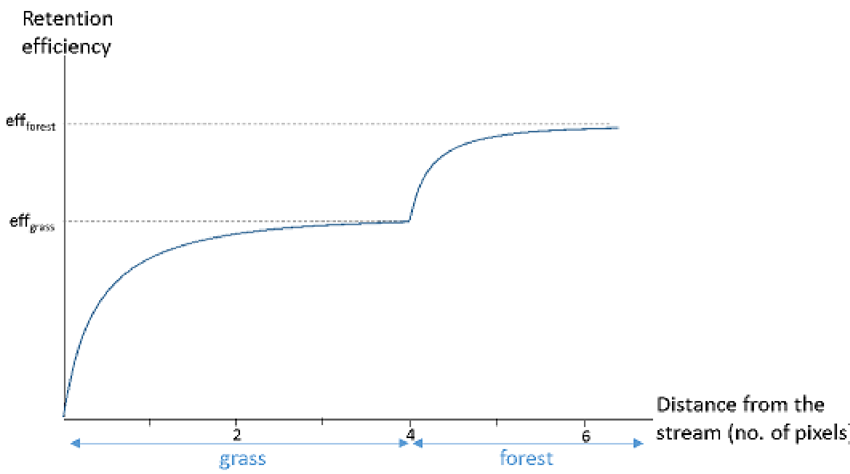


Fig. 5.3: Illustration of the calculation of the retention efficiency along a simple flow path composed of 4 pixels of grass and 3 pixels of forest. Each additional pixel of the grass LULC contributes to a smaller percentage toward the maximum efficiency provided by grass. The shape of the exponential curves is determined by the maximum efficiency and the retention length.

IC, the index of connectivity, represents the hydrological connectivity, i.e. how likely nutrient on a pixel is likely to reach the stream. In this model, IC is a function of topography only (Figure 3):

$$IC = \log_{10}\left(\frac{D_{up}}{D_{dn}}\right) \tag{5.23}$$

where

- $D_{up} = \overline{S}\sqrt{A}$ and,
- $D_{dn} = \sum_i \frac{d_i}{S_i}$

where :math:'D_{up} = overline{S} is the average slope gradient of the upslope contributing area (m/m), A is the upslope contributing area (m2); di is the length of the flow path along the ith cell according to the steepest downslope direction (m) (see details in sediment model), and Si is the slope gradient of the ith cell, respectively.

Note: The upslope contributing area and downslope flow path are delineated with the D-infinity flow algorithm (Tarboton, 1997). To avoid infinite values for IC, slope values S are forced to a minimum of 0.005 m/m if they occur to be less than this threshold, based on the DEM (Cavalli et al., 2013).

The value of IC_0 is set to $IC_0 = \frac{IC_{max}+IC_{min}}{2}$. This imposes that the sigmoid function relating NDR to IC is centered on the median of the IC distribution, hence that the maximum IC value gives $NDR = NDR_{max}$. k is set to a default value of 2 (cf. SDR model theory); it is an empirical factor that represents local topography.

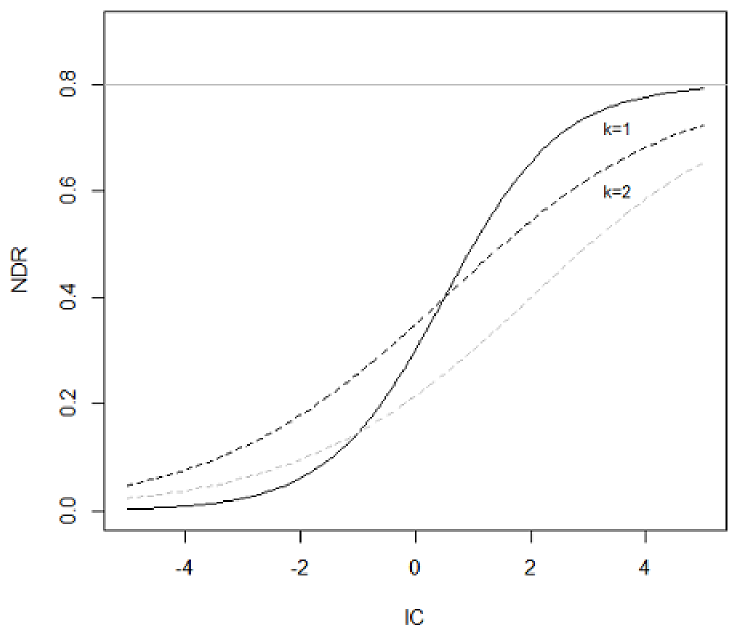


Fig. 5.4: Relationship between NDR and the connectivity index IC. The maximum value of NDR is set to $NDR_0 = 0.8$. The effect of the calibration is illustrated by setting $k = 1$ and $k = 2$ (solid and dashed line, respectively), and $IC_0 = 0.5$ and $IC_0 = 2$ (black and gray dashed lines, respectively).

Subsurface NDR

The expression for the subsurface NDR is a simple exponential decay with distance to stream, plateauing at the value corresponding to the user-defined maximum subsurface nutrient retention:

$$NDR_{subs,i} = 1 - eff_{subs}\left(1 - e^{\frac{-5\cdot\ell}{\ell_{subs}}}\right) \tag{5.24}$$

where

- eff_{subs} is the maximum nutrient retention efficiency that can be reached through subsurface flow (i.e. retention due to biochemical degradation in soils),
- ℓ_{subs} is the subsurface flow retention length, i.e. the distance after which it can be assumed that soil retains nutrient at its maximum capacity,
- ℓ_i is the distance from the pixel to the stream.

Nutrient export

Nutrient export from each pixel i is calculated as the product of the load and the NDR:

$$x_{exp_i} = load_{surf,i}\cdot NDR_{surf,i} + load_{subs,i}\cdot NDR_{subs,i} \tag{5.25}$$

Total nutrient at the outlet of each user-defined watershed is the sum of the contributions from all pixels within that watershed:

$$x_{exp_{tot}} = \sum_i x_{exp_i} \tag{5.26}$$

Limitations

The model has a small number of parameters and outputs generally show a high sensitivity to inputs. This implies that errors in the empirical load parameter values will have a large effect on predictions. Similarly, the retention efficiency values are based on empirical studies, and factors affecting these values (like slope or intra-annual variability) are averaged. These values implicitly incorporate information about the dominant nutrient dynamics, influenced by climate and soils. Finally, the effect of grid resolution on the NDR formulation has not been well studied.

Sensitivity analyses are recommended to investigate how the confidence intervals in input parameters affect the study conclusions (Hamel et al., 2015).

Also see the "Biophysical model interpretation" section for further details on model uncertainties.

Options for Valuation

Nutrient export predictions can be used for quantitative valuation of the nutrient retention service. For example, scenario comparison can serve to evaluate the change in purification service between landscapes. The total nutrient load can be used as a reference point, assuming that the landscape has 0 retention. Comparing the current scenario export to the total nutrient load provides a quantitative measure of the retention service of the current landscape.

Data Needs

This section outlines the data used by the model. Refer to the appendix for detailed information on data sources and pre-processing. For all raster inputs, the projection should be defined and the projection's linear units should be in meters.

1. **Digital elevation model** (DEM) (required). A GIS raster dataset, with an elevation value for each cell. Make sure the DEM is corrected by filling in sinks. To ensure proper flow routing, the DEM should extend beyond the watersheds of interest, rather than being clipped to the watershed boundaries. See the Working with the DEM section of this manual for more information.
2. **Land use/land cover** (required). A GIS raster dataset, with an integer LULC code for each pixel. The LULC code should be an integer.
3. **Watersheds** (required). A shapefile of polygons. This is a layer of watersheds such that each watershed contributes to a point of interest where water quality will be analyzed. See the Working with the DEM section for information on creating watersheds.
4. **Biophysical Table** (required). A .csv table of land use/land cover (LULC) classes, containing data on water quality coefficients used in this tool. These data are attributes of each LULC class rather than attributes of individual cells in the raster map. Each row in the table is an LULC class while each column contains a different attribute of each land use/land cover class. The columns must be named as:

   - *lucode* (Land use code): Unique integer for each LULC class (e.g., 1 for forest, 3 for grassland, etc.), must match the LULC raster above.
   - *LULC_desc*: Descriptive name of land use/land cover class (optional)
   - *load_n* (and/or load_p): The nutrient loading for each land use, given as decimal values with units of kg. ha-1 yr -1. Suffix _n stands for nitrogen, and _p for phosphorus, and the two compounds can be modeled at the same time or separately.

   Note 1: Loads are the sources of nutrients associated with each LULC. If the user wants to represent different level of fertilizer application, he/she needs to create different LULC.

   Note 2: Load values may be expressed either as the amount of nutrient applied (e.g. fertilizer, livestock waste, atmospheric deposition); or as “extensive” measures of contaminants, which are empirical values representing the contribution of a parcel to the nutrient budget (e.g. nutrient export running off urban areas, crops, etc.) In the latter case, the load should be corrected for the nutrient retention from downstream pixels of the same LULC. For example, if the measured (or empirically derived) export value for forest is 3 kg.ha-1.yr-1 and the retention efficiency is 0.8, users should enter 15(kg.ha-1.yr-1) in the n_load column of the biophysical table; the model will calculate the nutrient running off the forest pixel (n_export) as 15*0.8 = 3 kg.ha-1.yr-1.

   - *eff_n* (and/or eff_p): The maximum retention efficiency for each LULC class, varying between zero and 1. The nutrient retention capacity for a given vegetation is expressed as a proportion of the amount of nutrient from upstream. For example, high values (0.6 to 0.8) may be assigned to all natural vegetation types (such as forests, natural pastures, wetlands, or prairie), indicating that 60-80% of nutrient is retained. Like above, suffix _n stands for nitrogen, and _p for phosphorus, and the two compounds can be modeled at the same time or separately.
   - *ret_len_n* (and/or ret_len_p) (in meter): the distance after which it is assumed that a patch of LULC retains nutrient at its maximum capacity. If nutrients travel a distance smaller than the retention length, the retention efficiency will be less than the maximum value eff_x, following an exponential decay (see Nutrient transport section)
   - *proportion_subsurface_n* (optional): the proportion of dissolved nutrients over the total amount of nutrients, expressed as ratio between 0 and 1. By default, this value should be set to 0, indicating that all nutrients are delivered via surface flow.

   Example:

lucode	LULC_desc	load_n	eff_n	crit_len_n	proportion_subsurface_n
1	Agriculture	100	0.5	25	0.5
2	Grass	8	0.75	150	0
3	Forest	2.8	0.8	300	0
4	Wetland	2.8	0.8	10	0
5	Urban	10	0.05	10	0

5. **Subsurface_retention_efficiency (Nitrogen or phosphorus)**: the maximum nutrient retention efficiency that can be reached through subsurface flow, a value between 0 and 1. This field characterizes the retention due to biochemical degradation in soils.

6. **Subsurface_ret_len (Nitrogen or phosphorus)** (in meter): the distance (traveled subsurface and downslope) after which it is assumed that soil retains nutrient at its maximum capacity. If dissolved nutrients travel a distance smaller than subsubsurface_ret_len, the retention efficiency is lower than the maximum value defined above. Setting this value to a distance smaller than the pixel size will result in the maximum retention efficiency being reached within one pixel only.

7. **Threshold flow accumulation value**: Integer value defining the number of upstream pixels that must flow into a pixel before it's considered part of a stream. This is used to generate a stream layer from the DEM (see RouteDEM documentation of the InVEST manual). This threshold expresses where hydrologic routing is discontinued, i.e. where retention stops and the remaining pollutant will be exported to the stream. The default is 1 over the pixel area (in km2), i.e. ~1000 for 30m resolution. If the user has a map of stream lines in the watershed of interest, he/she should "calibrate" the threshold value by comparing the map with the *stream.tif* map output by the model.

8. **Borselli k parameter**: calibration parameter that determine the shape of the relationship between hydrologic connectivity (the degree of connection from patches of land to the stream) and the sediment delivery ratio (percentage of soil loss that actually reaches the stream; cf. Figure 2). The default value is 2.

5.5.4 Running the Model

To launch the nutrient model navigate to the Windows Start Menu -> All Programs -> InVEST +VERSION+ -> Nutrient delivery and retention. The interface does not require a GIS desktop, although the results will need to be explored with any GIS tool including ArcGIS, QGIS, and others.

Interpreting results

Model outputs

The following is a short description of each of the outputs from the standalone Nutrient Delivery and retention model. These results are found within the model's workspace specified in the user interface.

- **Parameter log**: Each time the model is run, a text (.txt) file will appear in the *Output* folder. The file will list the parameter values for that run and will be named according to the service, the date and time, and the suffix.
- **Output folder**:
  - **outputx_export_suffix.shp**: This is a shapefile which aggregates the nutrient model results per watershed, with x being n for nitrogen, and p for phosphorus. The .dbf table contains the following information for each watershed:
    - *x_load_tot*: $\mathrm{kg.yr}^{-1}$: total nutrient loads (sources) in the watershed, i.e. the sum of the nutrient contribution from all LULC without filtering from the landscape.
    - *x_exp_tot*: $\mathrm{kg.yr}^{-1}$: total nutrient export from the watershed

- **outputx_export.tif** : (kg/pixel) A pixel level map showing how much load from each pixel eventually reaches the stream.

- **Intermediate folder**:
  - *ret_len_x*: map of retention length values, ret_len, found in the biophysical table
  - d_dn: downslope factor of the index of connectivity (Eq. 5)
  - *d_up*: distance from a pixel to the stream (following the D-infinity algorithm, see RouteDEM documentation for details)
  - *d_up*: map of the retention efficiencies, eff_x, found in the biophysical table
  - *effective_retention_x*: map of the effective retention provided by the downslope flow path for each pixel (Eq. 3)
  - *ic_factor*: map of the index of connectivity (Eq. 5)
  - *load_n*: map of loads (for surface transport) per pixel (kg,yr-1)
  - *ndr_x*: map of NDR values
  - *s_accumulation.s_bar*: slope parameters for IC equation found in the Nutrient transport section
  - *stream*: stream network computed by the RouteDEM algorithm (with 0s representing land pixels, and 1s representing stream pixels)
  - *sub_crit_len_x*: map of the critical distance value for subsurface transport, subsurface_crit_len_x (constant over the landscape)
  - ***sub_eff_x*: map of the subsurface retention efficiency, subsurface_retention_eff (constant over the landscape)**

    - *sub_effective_retention_x: map of the subsurface effective retention (Eq. 7)
  - *sub_load_x*: map of nutrient loads for subsurface transport, per pixel (kg,yr-1)
  - *sub_ndr_x*: map of subsurface NDR values

- Prepared_data folder: Contains low-level hydrological routing outputs from the RouteDEM module including flow direction, flow accumulation, and slope.

Biophysical Model Interpretation for Valuation

Some valuation approaches, e.g. those relying on the changes in water quality for a treatment plant, are very sensitive to the model absolute predictions. Therefore, it is important to consider the uncertainties associated with the use of InVEST as a predictive tool and minimize their effect on the valuation step.

Model parameter uncertainties

Uncertainties in input parameters can be characterized during the literature review (e.g. examining the distribution of values from different studies). One option to assess the impact of parameter uncertainties is to conduct local or global sensitivity analyses, with the ranges obtained from the literature (Hamel et al., 2015).

Model structural uncertainties

The InVEST model computes a nutrient mass balance over a watershed, subtracting nutrient losses (conceptually represented by the retention coefficients), from the total nutrient sources. Where relevant, it is possible to distinguish between surface and subsurface flow paths, adding three parameters to the model. In the absence of empirical knowledge, modelers can assume that the surface load and retention parameters represent both transport process. Testing and calibration of the model is encouraged, acknowledging the main two challenges:

- knowledge gaps in nutrient transport: although there is strong evidence of the impact of land use change on nutrient export, modeling of the watershed scale dynamics remains challenging (Breuer et al., 2008; Scanlon et al., 2007). Calibration is therefore difficult and not recommended without in-depth analyses that would provide confidence in model process representation (Hamel et al., 2015)
- Potential contribution from point source pollution: domestic and industrial waste are often part of the nutrient budget and should be accounted for during calibration (for example, by adding point-source nutrient loads to modeled nutrient export, then comparing the sum to observed data).

Comparison to observed data

Despite the above uncertainties, the InVEST model provides a first-order assessment of the processes of nutrient retention and may be compared with observations. Time series of nutrient concentration used for model validation should span over a reasonably long period to attenuate the effect of interannual variability. Time series should also be relatively complete throughout a year (without significant seasonal data gaps) to ensure comparison with total annual loads. If the observed data is expressed as a time series of nutrient concentration, they need to be converted to annual loads (LOADEST and FLUX32 are two software facilitating this conversion). Additional details on methods and model performance for relative predictions can be found in the study of Hamel et al. (in prep).

5.5.5 Appendix: Data sources

This is a non-exhaustive list of data sources and suggestions about finding, compiling, and formatting data. It is updated as new data sources and methods become available.

In general, the FAO Geonetwork can be a valuable data source for different GIS layers for users outside the United States: http://www.fao.org/geonetwork/srv/en/main.home.

1. **Digital elevation model** (DEM)

   DEM data is available for any area of the world, although at varying resolutions. A list of free global DEMs are available at http://vterrain.org/Elevation/global.html.

   Free raw global DEM data is available from:

   - the World Wildlife Fund - http://worldwildlife.org/pages/hydrosheds
   - NASA: http://asterweb.jpl.nasa.gov/gdem-wist.asp (30m resolution)
   - USGS: http://eros.usgs.gov/elevation-products and http://hydrosheds.cr.usgs.gov/.

   Alternatively, it may be purchased relatively inexpensively at sites such as MapMart (www.mapmart.com).

   The DEM resolution may be a very important parameter depending on the project's goals. For example, if decision makers need information about impacts of roads on ecosystem services then fine resolution is needed. The hydrological aspects of the DEM used in the model must be correct. Because the model requires that all pixels have a flow direction (according to the D-infinity flow algorithm (Tarboton, 1997)), the DEM may need to be filled to remove sinks. Multiple passes of the ArcGis Fill tool, or Qgis Wang&Liu Fill algorithm (SAGA library) have shown good results.

2. Land use and land cover

   A key component for all water models is a spatially continuous landuse / land cover raster grid. That is, within a watershed, all landuse / land cover categories should be defined. Gaps in data will create errors. Unknown data gaps should be approximated. Global land use data is available from:

- the University of Maryland's Global Land Cover Facility: http://glcf.umd.edu/data/landcover/ (data available in 1 degree, 8km and 1km resolutions).
- NASA: https://lpdaac.usgs.gov/products/modis_products_table/mcd12q1 (MODIS multi-year global landcover data provided in several classifications)
- the European Space Agency: http://due.esrin.esa.int/globcover/ (landcover maps for 2005 and 2009)

Data for the U.S. for 1992, 2001 and 2011 is available as the National Land Cover Data product, produced by the Multi-Resolution Land Characteristics (MRLC) Consortium (a partnership of federal agencies): http://www.mrlc.gov

The simplest categorization of LULCs on the landscape involves delineation by land cover only (e.g., cropland, temperate conifer forest, prairie). Several global and regional land cover classifications are available (e.g., Anderson et al. 1976), and often detailed land cover classification has been done for the landscape of interest.

A slightly more sophisticated LULC classification involves breaking relevant LULC types into more model-relevant types. For example, agricultural land classes could be broken up into different crop types or forest could be broken up into specific species. The categorization of land use types depends on the model and how much data is available for each of the land types. Users should only break up a land use type if it will provide more accuracy in modeling. For instance, for the sediment model the user should only break up "crops" into different crop types if they have information on the difference in soil characteristics between crop management values.

The categorization of land use types depends on the model and how much data is available for each of the land types. The user should only break up a land use type if it will provide more accuracy in modeling. For instance, for the Nutrient delivery and Retention model the user should only break up 'crops' into different crop types if they have information on the difference in nutrient loading between crops. Along the same lines, the user should only break the forest land type into specific species for the water supply model if information is available on the root depth and evapotranspiration coefficients for the different species.

3. Watersheds / subwatersheds

   Watersheds outlets should correspond to reservoirs or other points of interest. This ensures that the sediment loads predicted by the model can be compared to observed data at these points. If known watershed maps exist, they should be used. Otherwise, watersheds and subwatersheds can be generated in ArcMap or QGIS based on the digital elevation model (see section on DEM for use of Fill tools to correct flow paths).

   Exact locations of specific structures, such as reservoirs, should be obtained from the managing entity or may be obtained on the web at sites such as the National Inventory of Dams (http://geo.usace.army.mil/pgis/f?p=397:1:0). Global collections of dam locations and information include the Global Reservoir and Dam (GRanD) Database (http://www.gwsp.org/products/grand-database.html) and the World Water Development Report II dam database (http://wwdrii.sr.unh.edu/download.html.)

4. Nutrient load parameter

   For all water quality parameter (nutrient load, retention efficiency, and retention length), local literature should be consulted to derive site-specific values. The NatCap database provides a non-exhaustive list of local references for nutrient loads and retention efficiencies. Parn et al. (2012) and Harmel et al. (2007) provide a good review for agricultural land in temperate climate.

   Examples of export coefficients ("extensive" measures, see Data needs) for the US can be found in the EPA PLOAD User's Manual and in a review by Lin (2004)[http://el.erdc.usace.army.mil/elpubs/pdf/tnwrap04-

3.pdf]. Note that the examples in the EPA guide are in lbs/ac/yr and would need to be converted to kg/ha/yr.

5. Retention efficiency

   This value represents, conceptually, the maximum nutrient retention that can be expected from a given LULC. Natural vegetation LULC types (such as forests, natural pastures, wetlands, or prairie) are assigned high values (>0.8). A review of the local literature and consultation with hydrologists is recommended to select the most relevant values for this parameter. Parn et al. provide a useful review for temperate climates. Reviews of riparian buffers efficiency, although a particular case of LULC retention, can also be used as a starting point (Mayer et al., 2007; Zhang et al., 2009).

6. Retention length

   This value represents the typical distance necessary to reach the maximum retention efficiency. It was introduced in the model to remove any sensitivity to the resolution of the LULC raster. The literature on riparian buffer removal efficiency suggests that retention lengths range from 10 to 300 m (Mayer et al., 2007; Zhang et al., 2009). In the absence of local data for land uses that are not forest or grass, one can simply set the retention length constant, equal to the pixel size: this will result in the maximum retention efficiency being reached within a distance of one pixel only.

7. Subsurface parameters: proportion_subsurface_n, eff_sub, crit_len_sub

   These values are used for advanced analyses and should be selected in consultation with hydrologists. Parn et al. (2012) provide average values for the partitioning of N loads between leaching and surface runoff. From Mayer et al. (2007), a global average of 200m for the retention length, and 80% for retention efficiency can be assumed for vegetated buffers.

5.5.6 References

Breuer, L., Vaché, K.B., Julich, S., Frede, H.-G., 2008. Current concepts in nitrogen dynamics for mesoscale catchments. Hydrol. Sci. J. 53, 1059–1074.

California Regional Water Quality Control Board Central Coast Region, 2013. Total Maximum Daily Loads for Nitrogen Compounds and Orthophosphate for the Lower Salinas River and Reclamation Canal Basin , and the Moro Cojo Slough Subwatershed , Monterey County, CA. Appendix F. Available at: http://www.waterboards.ca.gov/centralcoast/water_issues/programs/tmdl/docs/salinas/nutrients/index.shtml

Endreny, T.A., Wood, E.F., 2003. Watershed weighting of export coefficients to map critical phosphorous loading areas. J. Am. Water Resour. Assoc. 08544, 165–181.

Hamel, P., Chaplin-Kramer, R., Sim, S., Mueller, C., 2015. A new approach to modeling the sediment retention service (InVEST 3.0): Case study of the Cape Fear catchment, North Carolina, USA. Sci. Total Environ. 166–177.

Harmel, D., Potter, S., Casebolt, P., Reckhow, K., 2007. Compilation of measured nutrient load data for agricultural land uses in the United States 76502, 1163–1178.

Heathwaite, A.L., Quinn, P.F., Hewett, C.J.M., 2005. Modelling and managing critical source areas of diffuse pollution from agricultural land using flow connectivity simulation. J. Hydrol. 304, 446–461.

Keeler, B.L., Polasky, S., Brauman, K.A., Johnson, K.A., Finlay, J.C., Neill, A.O., 2012. Linking water quality and well-being for improved assessment and valuation of ecosystem services 109, 18629–18624.

Lin, J.., 2004. Review of published export coefficient and event mean concentration (EMC) data, WRAP Technical Notes Collection (ERDC TN-WRAP-04-3). Vicksburg, MS.

Mayer, P.M., Reynolds, S.K., Mccutchen, M.D., Canfield, T.J., 2007. Meta-Analysis of Nitrogen Removal in Riparian Buffers 1172–1180.

Pärn, J., Pinay, G., Mander, Ü., 2012. Indicators of nutrients transport from agricultural catchments under temperate climate: A review. Ecol. Indic. 22, 4–15.

Reckhow, K.H., Beaulac, M.N., Simpson, J.T., 1980. Modeling Phosphorus loading and lake response under uncertainty: A manual and compilation of export coefficients. EPA 440/5-80-011. US-EPA, Washington, DC.

Scanlon, B.R., Jolly, I., Sophocleous, M., Zhang, L., 2007. Global impacts of conversions from natural to agricultural ecosystems on water resources: Quantity versus quality. Water Resour. Res. 43.

Tarboton, D., 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. Water Resour. Res. 33, 309–319.

Vigiak, O., Borselli, L., Newham, L.T.H., Mcinnes, J., Roberts, A.M., 2012. Comparison of conceptual landscape metrics to define hillslope-scale sediment delivery ratio. Geomorphology 138, 74–88.

Zhang, X., Liu, X., Zhang, M., Dahlgren, R. a, Eitzel, M., 2009. A review of vegetated buffers and a meta-analysis of their mitigation efficacy in reducing nonpoint source pollution. J. Environ. Qual. 39, 76–84.

5.6 Sediment Delivery Ratio model

5.6.1 Summary

The objective of the InVEST sediment delivery model is to map overland sediment generation and delivery to the stream. In a context of global change, such information can be used to study the service of sediment retention in a catchment. This is of particular interest for reservoir management and instream water quality, both of which may be economically valued.

The main differences between the InVEST SDR model and the InVEST Sediment retention model found in InVEST v3.0.1 and earlier are:

- The routing of sediment from a cell to the stream was modified to remove the sensitivity to grid resolution and facilitate the selection of LULC-specific retention coefficient;
- (Optional) calibration is based on one non-physical parameter that preserves the spatial distribution of sediment sinks and sources, facilitating the interpretation of spatially explicit outputs;
- The increased flexibility in model structure allows advanced users to represent more complex processes such as gully erosion or instream retention (work is in progress to facilitate the representation of these processes for InVEST users)

5.6.2 Introduction

Erosion and overland sediment retention are natural processes that govern the sediment concentration in streams. Sediment dynamics at the catchment scale are mainly determined by climate (in particular the rain intensity), soil properties, topography, and vegetation; and anthropogenic factors such as agricultural activities or dam construction and operation. Main sediment sources include overland erosion (soil particles detached and transported by rain and overland flow), gullies (channels that concentrate flow), bank erosion, and mass erosion (or landslides; see Merrit 2003 for a review). Sinks include on-slope, floodplain or instream deposition, and reservoir retention, as summarized in Figure 1. Conversion of land use and changes in land management practices may dramatically modify the amount of sediment running off a catchment. The magnitude of this effect is primarily governed by: i) the main sediment sources (land use change will have a smaller effect in catchments where sediments are not primarily coming from overland flow); and ii) the spatial distribution of sediment sources and sinks (for example, land use change will have a smaller effect if the sediment sources are buffered by vegetation).

Increases in sediment yield are observed in many places of the world, dramatically affecting water quality and reservoir management (UNESCO 2009). The sediment retention service provided by natural landscapes is of great interest to water managers. Understanding where the sediments are produced and delivered allow them to design improved strategies for reducing sediment loads. Changes in sediment load can have impacts on downstream irrigation, water treatment, recreation and reservoir performance.

The outputs from the sediment model include the sediment load delivered to the stream at an annual time scale, as well as the amount of sediment eroded in the catchment and retained by vegetation and topographic features. The model provides two options for valuation of the sediment retention service, though appropriate valuation approaches will be highly dependent on the particular application and context, and may need to be implemented independent of InVEST.

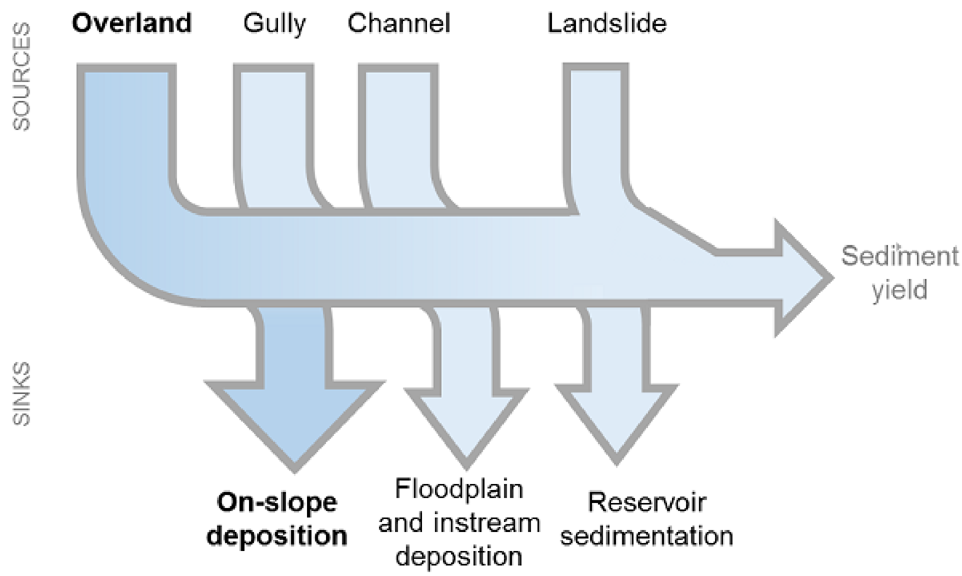


Fig. 5.5: General catchment sediment budget. The relative size of the arrows changes depending on the environment. The InVEST model focuses on the overland processes.

5.6.3 The Model

Biophysical Model

Sediment Delivery

The sediment delivery module is a spatially-explicit model working at the spatial resolution of the input DEM raster. For each cell, the model first computes the amount of eroded sediment, then the sediment delivery ratio (SDR), which is the proportion of soil loss actually reaching the catchment outlet. This approach was proposed by Borselli et al. (2008) and has received increasing interest in recent years (Cavalli et al., 2013; López-vicente et al., 2013; Sougnez et al., 2011). See Advantages and limitations for further discussion.

Annual Soil Loss The amount of annual soil loss on pixel i, $usle_i$ $(ton.ha^{-1}yr^{-1})$, is given by the revised universal soil loss equation (RUSLE1):

$$usle_i = R_i \cdot K_i \cdot LS_i \cdot C_i \cdot P_i,$$

where

- R_i is the rainfall erosivity $(MJ \cdot mm(ha \cdot hr)^{-1})$,
- K_i is the soil erodibility $(ton \cdot ha \cdot hr(MJ \cdot ha \cdot mm)^{-1})$,
- LS_i is the slope length-gradient factor
- C_i is the crop-management factor

- and P_i is the support practice factor (Renard et al., 1997). (cf. also in (Bhattarai and Dutta, 2006)).

and LS_i factor is given from the method developed by Desmet and Govers (1996) for two-dimension surface:

$$LS_i = S_i \frac{(A_{i-in} + D^2)^{m+1} - A_{i-in}^{m+1}}{D^{m+2} \cdot x_i^m \cdot (22.13)^m}$$

where

- S_i the slope factor for grid cell calculated as function of slope radians θ
- $S = 10.8 \cdot \sin(\theta) + 0.03$ where $\theta < 9\%$
- $S = 16.8 \cdot \sin(\theta) - 0.50$, where $\theta \geq 9\%$
- A_{i-in} the contributing area (m^2) at the inlet of a grid cell which is computed from the d-infinity flow direction method
- D the grid cell linear dimension (m)
- $x_i = |sin\alpha_i| + |\cos\alpha_i|$ where α_i is the aspect direction for grid cell i
- m is the RUSLE length exponent factor.

To avoid overestimation of the LS factor in heterogeneous landscapes, long slope lengths are capped to a value of 333m (Desmet and Govers, 1996; Renard et al., 1997)

The value of m, the length exponent of LS factor, is based on the classical USLE, as discussed in (Oliveira et al., 2013):

- $m = 0.2$ for slope <= 1%:
- $m = 0.3$ for 1% < slope <= 3.5%
- $m = 0.4$ for 3.5% < slope <= 5%
- $m = 0.5$ for 5% < slope <= 9%
- $m = \beta/(1+\beta)$ where $\beta = \sin\theta/0.0986/(3\sin\theta^{0.8} + 0.56)$ for slope >= 9%

Sediment Delivery Ratio **Step 1** Based on the work by Borselli et al. (2008), the model first computes the connectivity index:

$$IC = \log_{10}\left(\frac{D_{up}}{D_{dn}}\right)$$

Figure 2. Conceptual approach used in the model. The sediment delivery ratio (SDR) for each pixel is a function of the upslope area and downslope flow path (Equations 3, 4, 5).

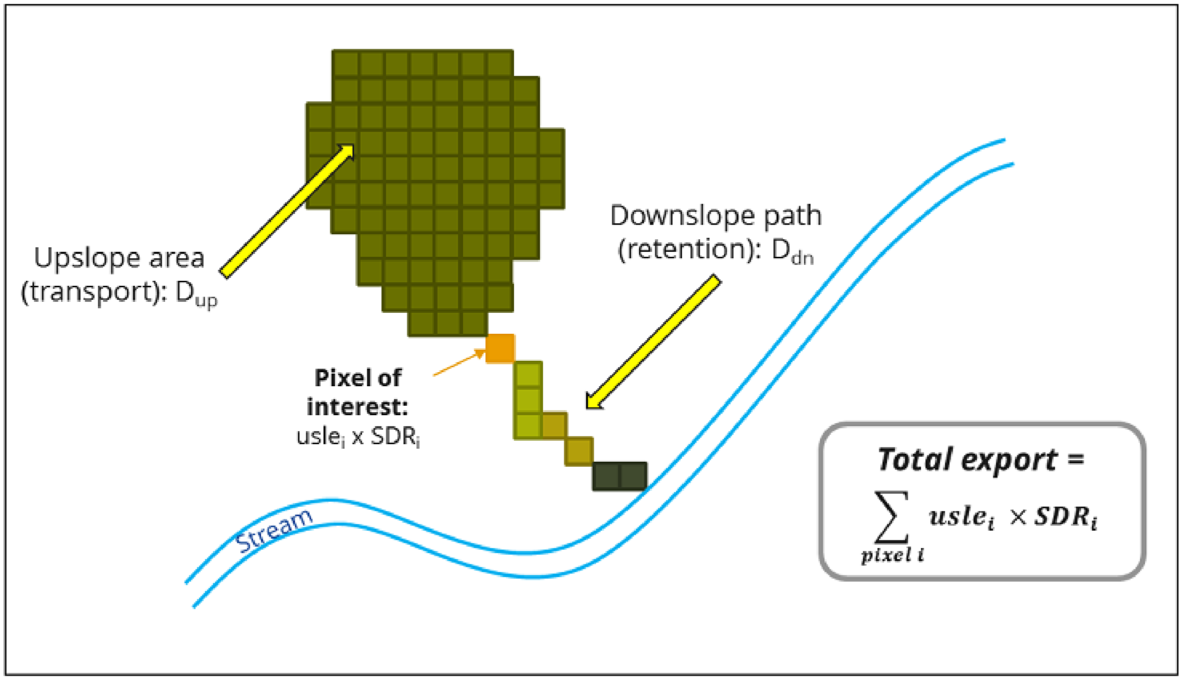
D_{up} is the upslope component defined as:

$$D_{up} = \bar{C}\bar{S}\sqrt{A}$$

where $\bar{C}$ is the average C factor of the upslope contributing area, S is the average slope gradient of the upslope contributing area (m/m) and A is the upslope contributing area (m^2). The upslope contributing area is delineated from the D-infinity flow algorithm (Tarboton, 1997).

The downslope component D_{dn} is given by:

$$D_{dn} = \sum_i \frac{d_i}{C_i S_i}$$



where d_i is the length of the flow path along the ith cell according to the steepest downslope direction (m) (see Figure 2), C_i and S_i are the C factor and the slope gradient of the ith cell, respectively. Again, the downslope flow path is determined from the D-infinity flow algorithm (Tarboton, 1997).

To avoid infinite values for IC, slope values S are forced to a minimum of 0.005 m/m if they occur to be less than this threshold, and an upper limit of 1 m/m to limit bias due to very high values of IC on steep slopes. (Cavalli et al., 2013).

Step 2 The SDR ratio for a pixel i is then derived from the conductivity index IC following (Vigiak et al., 2012):

$$SDR_i = \frac{SDR_{max}}{1 + \exp\left(\frac{IC_0 - IC_i}{k}\right)}$$

where SDR_{max} is the maximum theoretical SDR, set to an average value of 0.8 (Vigiak et al., 2012), and IC_0 and k are calibration parameters that define the shape of the SDR-IC relationship (increasing function). The effect of IC_0 and k on the SDR is illustrated below:

Figure 3. Relationship between the connectivity index IC and the SDR. The maximum value of SDR is set to $SDR_{max} = 0.8$. The effect of the calibration are illustrated by setting $k_b = 1$ and $k_b = 2$ (solid and dashed line, respectively), and $IC_0 = 0.5$ and $IC_0 = 2$ (black and grey dashed lines, respectively).

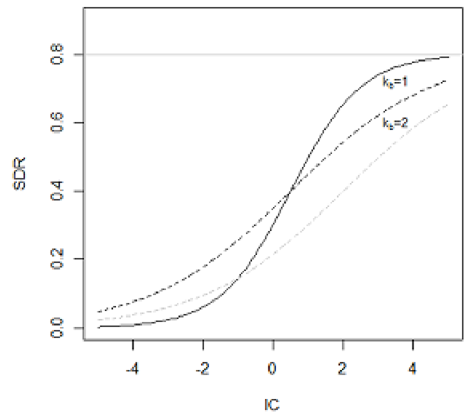
Sediment Load The sediment load from a given pixel i, E_i $(ton.ha^{-1}yr^{-1})$ is given by:

$$E_i = usle_i \cdot SDR_i$$

The total catchment sediment load E $(ton.ha^{-1}yr^{-1})$ is given by:

$$E = \sum_i E_i$$

E is the value used for calibration/validation purposes, in combination with other sediment sources, if data are available.



Optional Drainage Layer

In some situations, the index of connectivity defined by topography does not represent actual flow paths, which may be influenced by artificial connectivity instead. For example, sediments in urban areas or near roads are likely to be conveyed to the stream with little retention. The (optional) drainage raster identifies the pixels that are artificially connected to the stream, irrespective of their geographic position (e.g. their distance to the stream network). Pixels from the drainage layer are treated similarly to pixels of the stream network; in other words, the downstream flow path will stop at pixels of the drainage layer (and the corresponding sediment load will be added to the total sediment export).

Limitations to the Biophysical Model

- Among the main limitations of the model is its reliance on the USLE (Renard et al., 1997). This equation is widely used but is limited in scope, only representing rill/inter-rill erosion processes. Other sources of sediment include gully erosion, streambank erosion, and mass erosion. A good description of the gully and streambank erosion processes is provided by Wilkinson et al. 2014, with possible modeling approaches. Mass erosion is not represented in the model but can be a significant source in some areas (REF. India) or under certain land use change: http://water.epa.gov/scitech/datait/tools/warsss/box08.cfm
- A corollary is that the descriptions of the impact on ecosystem services (and any subsequent valuation) should account for the relative proportion of the sediment source from the model compared to the total sediment budget (see section on Evaluting sediment retention services)
- In addition, as an empirical equation developed in the United States, the USLE has shown limited performance in other areas – even when focusing on sheet and rill erosion (REF.) Based on local knowledge, users may modify the soil loss equation implemented in the model by altering the R, K, C, P inputs to reflect findings from local studies (Sougnez et al., 2011).
- The model is very sensitive to the k and IC0 parameters, which are not physically based. The emerging literature on the modeling approach used in the InVEST model (Cavalli et al., 2013; López-vicente et al., 2013; Sougnez et al., 2011; Vigiak et al., 2012) provides guidance to set this parameters, but users should be aware of this limitation when interpreting model absolute values.
- Given the simplicity of the model and low number of parameters, outputs are very sensitive to most input parameters. Errors in the empirical parameters of the USLE equations will therefore have a large effect on predictions. Sensitivity analyses are recommended to investigate how the confidence intervals in input parameters affect the study conclusions.

Differences between the InVEST v3.1 SDR model and the original approach developed by Borselli et al. (2008)

The InVEST SDR model is based on the concept of hydrological connectivity, as parameterized by Borselli et al. (2012). This approach was selected since it requires a minimal number of parameters, uses globally available data, and is spatially explicit. In a comparative study, Vigiak et al. (2012) suggested that the approach provides: "large improvement in predicting specific sediment yields, (ii) ease of implementation, (iii) scale-independency; and (iv) a formulation capable of accounting for landscape variables and topology in line with sedimentological connectivity concepts". The approach has also been used to predict the effect of land use change (Jamshidi et al., 2013). The following points summarize the differences between InVEST and the Borselli's model:

- The weighting factor is directly implemented as the USLE C factor (other researchers have used a different formulation, e.g. roughness index based on high-resolution DEM (Cavalli et al., 2013))
- The SDR_{max} parameter used by Borselli et al. is set to 0.8 by default to reduce the number of parameters. Vigiak et al. (2012) propose to define SDR max as the fraction of topsoil particles finer than coarse sand (<1 mm).

Evaluating Sediment Retention Services

Sediment Retention Services

Translating the biophysical impacts of altered sediment delivery to human well-being metrics depends very much on the decision context. Soil erosion, suspended sediment and deposited sediment can have both negative and positive impacts on various users in a watershed (Keeler et al, 2012). These include, but are not limited to:

- Reduced soil fertility to reduced water and nutrient holding capacity
- Increase in treatment costs for drinking water supply
- Reduced lake clarity diminishing the value of recreation
- Increase in total suspended solids impacting health and distribution of aquatic populations
- Increase in reservoir sedimentation diminishing reservoir performance or increasing sediment control costs
- Increase in harbor sedimentation requiring dredging to preserve harbor function

Sediment Retention Index

An index of sediment retention is computed by the model as follows:

$$R_i \cdot K_i \cdot LS_i(1 - C_iP_i)SDR_i$$

which represents the avoided soil loss by the current land use compared to bare soil, weighted by the SDR factor. This index underestimates retention since it does not account for the retention from upstream sediment flowing through the given pixel. Therefore, this index should not be interpreted quantitatively. We also note that in some situations, index values may be counter-intuitive: for example, urban pixels may have a higher index than forest pixels if they are highly connected to the stream. In other terms, the SDR (second factor) can be high for these pixels, compensating for a lower service of avoided soil loss (the first factor): this suggests that the urban environment is already providing a service of reduced soil loss compared to an area of bare soil.

Quantitative Valuation

Sediment retention at the subwatershed level From a valuation standpoint, an important metric is the difference in retention or yield across scenarios. For quantitative assessment of the retention service, the model uses as a benchmark a hypothetical scenario where all land is cleared to bare soil: the value of the retention service is then based on the difference between the sediment export from this bare soil catchment and that of the scenario of interest. This output is termed "sed_retention" in the watershed summary table.

Additional sources and sinks of sediment As noted in the model limitations, the omission of some sources and sinks of sediment (gully erosion, stream bank erosion, and mass erosion) should be considered in the valuation analyses. In some systems, these other sources of sediment may dominate and large changes in overland erosion may not make a difference to overall sediment concentrations in streams. In other words, if the sediment yields from two scenarios differ by 50%, and the part of rill/inter-rill erosion in the sediment budget in 60%, then the actual change valued for avoided reservoir sedimentation is 30%.

One complication when calculating the total sediment budget is that changes in climate or land use result in changes in peak flow during rain events, and are thus likely to affect the magnitude of gully and streambank erosion. While the magnitude of the change in other sediment sources is highly contextual it is likely to be in the same direction as the change in overland erosion: a higher sediment overland transport is indeed often associated with higher flows, which likely increase gully and bank erosion. Therefore, when comparing across scenarios, the absolute change may serve as a lower bound on the total impact of a particular climate or land use change.

Appendix 2 summarizes options to represent the additional sources and sinks in the model.

Replacement and avoided cost frameworks, versus willingness to pay approaches With many ecosystem service impacts, and sediment impacts in particular, the valuation is relatively simple if an avoided mitigation cost or replacement cost method is deemed appropriate. In this situation, beneficiaries are assumed to incur a cost that is a function of the biophysical metric (eg, suspended sediment increases treatment costs). However, it is important to recognize that the avoided cost or replacement cost approaches assume the mitigating actions are worthwhile for the actor undertaking them. For example, if a reservoir operator deems that the costs associated with dredging deposited sediment are not worth the benefits of regaining lost storage capacity, it is not appropriate to value all deposited sediment at the unit cost of dredging. Similarly, an increase in suspended sediment for drinking water supplies may be met by increasing treatment inputs or switching to an alternate treatment technology. Avoiding these extra costs could then be counted as economic benefits. However, in some contexts, private water users may decide that the increase in sediment content is acceptable, rather than incur additional treatment expenses. They are economically worse off, but by not paying for additional treatment, the replacement cost approach becomes an upper bound on their economic loss. Their economic loss is also no longer captured by their change in financial expenditures, which further complicates the analysis.

Note, however, this bounding approach may be entirely appropriate for initial assessment of the significance of different benefit streams i.e. if the most expensive approach does not have a significant impact, then there is no need to refine the analysis to utilize more detailed approaches such as willingness-to-pay (for consumers) or impacts on net revenues (for producers). However, if the impact is large and there is no good reason to believe that the relevant actors will undertake the mitigating activities, then a willingness-to-pay framework is the appropriate path to take. For an introduction to the techniques available, see http://ecosystemvaluation.org/dollar_based.htm.

Time considerations Generally, economic and financial analysis will utilize some form of discounting that recognizes the time value of money, benefits, and use of resources. Benefits and costs that accrue in the future "count for less" than benefits and costs that are born close to the present. It is important that any economic or financial analysis be cognizant of the fact that the SDR model represents only average annual impacts under steady state conditions. This has two implications for valuation. First, users must recognize that the impacts being valued may take some time to come about: It is not the case that the full steady state benefits would begin accruing immediately, even though many of the costs might be. Second, the annual averaging means that cost or benefit functions displaying nonlinearities on shorter timescales should (if possible) be transformed, or the InVEST output should be paired with other statistical analysis to represent important intra or interannual variability.

5.6.4 Data Needs

This section outlines the specific data used by the model. See the Appendix for detailed information on data sources and pre-processing. Note that all GIS inputs must be in the same projection and in linear meter units for accurate results.

1. **Digital elevation model (DEM)** (required). A GIS raster dataset with an elevation value for each cell. Make sure the DEM is corrected by filling in sinks, and if possible compare the output stream maps with hydrographic maps of the area. To ensure proper flow routing, the DEM should extend beyond the watersheds of interest, rather than being clipped to the watershed edge.
2. **Rainfall erosivity index (R)** (required). R is a GIS raster dataset, with an erosivity index value for each cell. This variable depends on the intensity and duration of rainfall in the area of interest. The greater the intensity and duration of the rain storm, the higher the erosion potential. The erosivity index is widely used, but in case of its absence, there are methods and equations to help generate a grid using climatic data. The units on the index values are $MJ \cdot mm \cdot (ha \cdot h \cdot yr)^{-1}$.
3. **Soil erodibility (K)** (required). K is a GIS raster dataset, with a soil erodibility value for each cell. Soil erodibility, K, is a measure of the susceptibility of soil particles to detachment and transport by rainfall and runoff. The units on the index values are $ton \cdot ha \cdot h \cdot (ha \cdot MJ \cdot mm)^{-1}$
4. **Land use/land cover (LULC)** (required). LULC is a GIS raster dataset, with an integer LULC code for each cell.
5. **Watersheds** (required). A shapefile of polygons. This is a layer of watersheds such that each watershed contributes to a point of interest where water quality will be analyzed.
6. **Biophysical table** (required). A .csv table containing model information corresponding to each of the land use classes. Each row is a land use/land cover class and columns should be named and defined as follows:
   (a) **lucode** (Land use code): Unique integer for each LULC class (e.g., 1 for forest, 3 for grassland, etc.), must match the LULC raster input.
   (b) **usle_c**: Cover-management factor for the USLE, a floating point value between 0 and 1.
   (c) **usle_p**: Support practice factor for the USLE, a floating point value between 0 and 1.
7. **Threshold flow accumulation** (required). The number of upstream cells that must flow into a cell before it is considered part of a stream, which is used to classify streams in the DEM. This threshold directly affects the expression of hydrologic connectivity and sediment export: when a flow path reaches the stream, sediment deposition stops and the sediment exported is assumed to reach the catchment outlet.
8. k_b and IC_0: two calibration parameters that determine the shape of the relationship between hydrologic connectivity (the degree of connection from patches of land to the stream) and the sediment delivery ratio (percentage of soil loss that actually reaches the stream; cf. Figure 3). The default values are $k_b = 2$ and $IC_0 = 0.5$.
9. $\mathbf{SDR_{max}}$: the maximum SDR that a pixel can reach, which is a function of the soil texture. More specifically, it is defined as the fraction of topsoil particles finer than coarse sand (1000 μm; Vigiak et al. 2012). This parameter can be used for calibration in advanced studies. Its default value is 0.8.
10. **Drainage layer (optional)** A raster with 0s and 1s, where 1s correspond to pixels artificially connected to the stream (by roads, stormwater pipes, etc.). The flow routing will stop at these "artificially connected" pixels, before reaching the stream network.

5.6.5 Running the Model

To launch the Sediment model navigate to the Windows Start Menu -> All Programs -> InVEST +VERSION+ -> Sediment delivery and retention. The interface does not require a GIS desktop, although the results will need to be explored with any GIS tool including ArcGIS, QGIS, and others.

Interpreting Results

The following is a short description of each of the outputs from the Sediment Retention model. Final results are found in the output folders within the user defined Workspace specified for this model.

- **[workspace]** folder:
  - **Parameter log**: Each time the model is run, a text (.txt) file will appear in the Output folder. The file will list the parameter values for that run and will be named according to the service, the date and time, and the suffix.
  - **rkls.tif** (tons/pixel): Total potential soil loss per pixel in the original land cover without the C or P factors applied from the RKLS equation, equivalent to the soil loss for bare soil.
  - **sed_export.tif** (tons/pixel): The total amount of sediment exported from each pixel that reaches the stream.
  - **stream.tif** (pixel mask): The pixel level mask of the calculated stream network, useful for interpreting pixel level output and checking the stream network computed by the model.
  - **stream.tif** (pixel mask): The pixel level mask of the calculated stream network, useful for interpreting pixel level output and checking the stream network computed by the model.
  - **stream_and_drainage.tif** (pixel mask): If a drainage layer is provided, this raster is the union of that layer with the calculated stream layer.
  - **usle.tif** (tons/pixel): Total potential soil loss per pixel in the original land cover calculated from the USLE equation.
  - **sed_retention.tif** (tons/pixel): Map of sediment retention with reference to a bare watershed.
  - **sed_retention_index.tif** (tons/pixel): Index of sediment retention, used to identified areas contributing more to retention with reference to a bare watershed. This is NOT the sediment retained on each pixel (see Section on the index in "Evaluating Sediment Retention Services" above).
  - **watershed_results_sdr.shp**: Table containing biophysical values for each watershed, with fields as follows:
  - **sed_export** (tons/watershed): Total amount of sediment exported to the stream per watershed. This should be compared to any observed sediment loading at the outlet of the watershed. Knowledge of the hydrologic regime in the watershed and the contribution of the sheetwash yield into total sediment yield help adjust and calibrate this model.
    - **usle_tot** (tons/watershed): Total amount of potential soil loss in each watershed calculated by the USLE equation.
    - **sed_retention** (tons/watershed): Difference in the amount of sediment delivered by the current watershed and a hypothetical watershed where all land use types have been cleared to bare soil.
- **[workspace]\intermediate_outputs** folder:
  - dem_offset, slope, thresholded_slope, flow_direction, flow_accumulation, stream: hydrologic rasters based on the DEM used for flow routing (outputs from RouteDEM, see corresponding chapter in User's Guide)
  - ls (and bare_soil)-> LS factor for USLE (Eq. 1 and 2)
  - w_bar (and bare_soil) -> mean weighting factor (C factor) for upslope contributing area (Eq. 4)
  - s_bar (and bare_soil) -> mean slope factor for upslope contributing area
  - d_up (and bare_soil) -> upslope factor of the index of connectivity (Eq. 4)
  - ws_factor (and bare_soil) -> denominator of the downslope factor (Eq. 5)
  - d_dn (and bare_soil) -> downslope factor of the index of connectivity (Eq. 5)
  - ic_factor (and bare_soil) -> index of connectivity (Eq. 3)
  - sdr_factor (and bare_soil) -> sediment delivery ratio (SDR; Eq. 6)

Comparison with Observations

The sediment yield (sed_export) predicted by the model can be compared with available observations. These can take the form of sediment accumulation in a reservoir or time series of Total Suspended Solids (TSS) or turbidity. In the former case, the units are the same as in the InVEST model (tons per year). For time series, concentration data need to be converted to annual loads (LOADEST and FLUX32 are two software facilitating this conversion). A global database of sediment yields for large rivers can be found on the FAO website: http://www.fao.org/nr/water/aquastat/sediment/index.stm Alternatively, for large catchments, global sediment models can be used to estimate the sediment yield. A review of such models was performed by de Vente et al. (2013).

Note when comparing with measured results that the SDR model A key thing to remember when comparing predictions to observations is that the model represents rill-inter-rill erosion only. As indicated in the Introduction three other sources of sediment may contribute to the sediment budget: gully erosion, stream bank erosion, and mass erosion. The relative importance of these processes in a given landscape needs to be determined to ensure adequate model interpretation.

5.6.6 Appendix 1: Data Sources

This section is a compilation of potential data sources and suggestions about finding, compiling, and formatting data. It is not an exhaustive list. Although we strive to update this section regularly with new data sources and methods, users are encouraged to seek local good quality data to improve the quality of model inputs.

Digital Elevation Model (DEM)

DEM data is available for any area of the world, although at varying resolutions.

Free raw global DEM data is available from:

- the World Wildlife Fund - http://worldwildlife.org/pages/hydrosheds
- NASA: http://asterweb.jpl.nasa.gov/gdem-wist.asp (30m resolution); and easy access to SRTM data: http://dwtkns.com/srtm/
- USGS: http://eros.usgs.gov/elevation-products and http://hydrosheds.cr.usgs.gov/

Alternatively, it may be purchased relatively inexpensively at sites such as MapMart (www.mapmart.com).

The DEM resolution may be a very important parameter depending on the project's goals. For example, if decision makers need information about impacts of roads on ecosystem services then fine resolution is needed. The hydrological aspects of the DEM used in the model must be correct. Because the model requires that all pixels have a flow direction (according to the D-infinity flow algorithm (Tarboton, 1997)), the DEM may need to be filled to remove sinks. Multiple passes of the ArcGis Fill tool, or Qgis Wang&Liu Fill algorithm (SAGA library) have shown good results.

Rainfall Erosivity Index (R)

R should be obtained from published values, as calculation is very tedious. For calculation, R equals the annual average of EI values, where E is the kinetic energy of rainfall (in $MJ \cdot ha^{-1}$) and I30 is the maximum intensity of rain in 30 minutes (in mm.hr-1). A review of relationships between precipitation and erosivity index around the world is provided by Renard and Freimund (1994).

General guidance to calculate the R index can be found in the FAO Soils bulletin 70 (Roose, 1996): http://www.fao.org/docrep/t1765e/t1765e0e.htm

In the United States, national maps of the erosivity index can be found through the United States Department of Agriculture (USDA) and Environmental Protection Agency (EPA) websites. The USDA published a loss handbook (http://www.epa.gov/npdes/pubs/ruslech2.pdf) that contains a hard copy map of the erosivity index for each region.

Using these maps requires creating a new line feature class in GIS and converting to raster. Please note that conversion of units is also required: multiplication by 17.02 is needed to convert from US customary units to MJ.mm.(ha.h.yr)-1, as detailed in Appendix A of the USDA RUSLE handbook (Renard et al., 1997).

The EPA has created a digital map that is available at http://www.epa.gov/esd/land-sci/emap_west_browser/pages/wemap_mm_sl_rusle_r_qt.htm . The map is in a shapefile format that needs to be converted to raster, along with an adjustment in units.

Soil Erodibility (K)

Texture is the principal factor affecting K, but soil profile, organic matter and permeability also contribute. It varies from 70/100 for the most fragile soil and 1/100 for the most stable soil (in US customary units). Erodibility is typically measured on bare reference plots, 22.2 m-long on 9% slopes, tilled in the direction of the slope and having received no organic matter for three years.

The FAO provides global soil data in their Harmonized World Soil Database: http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/.

Soil data for many parts of the world are also available from the Soil and Terrain Database (SOTER) Programme (http://www.isric.org/projects/soil-and-terrain-database-soter-programme). Or in te

In the United States free soil data is available from the U.S. Department of Agriculture's NRCS in the form of two datasets:

SSURGO http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs142p2_053627 and

STATSGO http://water.usgs.gov/GIS/metadata/usgswrd/XML/ussoils.xml.Where available SSURGO data should be used, as it is much more detailed than STATSGO. Where gaps occur in the SSURGO data, STATSGO can be used to fill in the blanks. The Soil Data Viewer (http://www.nrcs.usda.gov/wps/portal/nrcs/detailfull/soils/home/?cid=nrcs142p2_053620) helps with pre-processing and downloading of the data.

Please note that conversion of units may be required: multiplication by 0.1317 is needed to convert from US customary units to $ton \cdot ha \cdot hr \cdot (ha \cdot MJ \cdot mm)^{-1}$, as detailed in Appendix A of the USDA RUSLE handbook (Renard et al., 1997).

Alternatively, the following equation can be used to calculate K (Renard et al., 1997):

$$K = \frac{2.1 \cdot 10^{-4}(12-a)M^{1.14} + 3.25(b-2) + 2.5(c-3)}{759}$$

In which K = soil erodibility factor $(t \cdot ha \cdot hr \cdot (MJ \cdot mm \cdot ha)^{-1}$; M = (silt $(\%)$ + very fine sand $(\%)$)(100-clay $(\%)$) a = organic matter $(\%)$ b = structure code: (1) very structured or particulate, (2) fairly structured, (3) slightly structured and (4) solid c = profile permeability code: (1) rapid, (2) moderate to rapid, (3) moderate, (4) moderate to slow, (5) slow and (6) very slow.

When profile permeability and structure are not available, soil erodibility can be estimated based on soil texture and organic matter content, based on the work of Wischmeier, Johnson and Cross (reported in Roose, 1996). The OMAFRA fact sheet summarize these values in the following table (http://www.omafra.gov.on.ca/english/engineer/facts/12-051.pdf):

	Average OMC	OMC<2%	OMC>2%
Clay	0.22	0.24	0.21
Clay loam	0.3	0.33	0.28
Coarse sandy loam	0.07	0.07	0.07
Fine sand	0.08	0.09	0.06
Fine sandy loam	0.18	0.22	0.17
Heavy clay	0.17	0.19	0.15
Loam	0.3	0.34	0.26
Loamy fine sand	0.11	0.15	0.09
Loamy sand	0.04	0.05	0.04
Loamy very fine sand	0.39	0.44	0.25
Sand	0.02	0.03	0.01
Sandy clay loam	0.2	0.2	0.2
Sandy loam	0.13	0.14	0.12
Silt loam	0.38	0.41	0.37
Silty clay	0.26	0.27	0.26
Silty clay loam	0.32	0.35	0.3
Very fine sand	0.43	0.46	0.37
Very fine sandy loam	0.35	0.41	0.33

Soil erodibility values (K) in US customary units based on the OMAFRA Fact sheet. Soil textural classes can be derived from the FAO guidelines for soil description (FAO, 2006, Figure 4).

A particular case is the K value for water bodies, for which soil maps may not indicate any soil type. A value of 0 can be substituted, assuming that no soil loss occurs in water bodies.

Land Use/Land Cover

A key component for all water models is a spatially continuous landuse / land cover raster grid. That is, within a watershed, all landuse / land cover categories should be defined. Gaps in data will create errors. Unknown data gaps should be approximated. Global land use data is available from:

- the University of Maryland's Global Land Cover Facility: http://glcf.umd.edu/data/landcover/ (data available in 1 degree, 8km and 1km resolutions).
- NASA: https://lpdaac.usgs.gov/products/modis_products_table/mcd12q1 (MODIS multi-year global landcover data provided in several classifications)
- the European Space Agency: http://due.esrin.esa.int/globcover/ (landcover maps for 2005 and 2009)

Data for the U.S. for 1992 and 2001 is provided by the EPA in their National Land Cover Data product: http://www.epa.gov/mrlc/.

The simplest categorization of LULCs on the landscape involves delineation by land cover only (e.g., cropland, temperate conifer forest, prairie). Several global and regional land cover classifications are available (e.g., Anderson et al. 1976), and often detailed land cover classification has been done for the landscape of interest.

A slightly more sophisticated LULC classification involves breaking relevant LULC types into more meaningful types. For example, agricultural land classes could be broken up into different crop types or forest could be broken up into specific species. The categorization of land use types depends on the model and how much data is available for each of the land types. Users should only break up a land use type if it will provide more accuracy in modeling. For instance, for the sediment model the user should only break up 'crops' into different crop types if they have information on the difference in soil characteristics between crop management values.

P and C Coefficients

The support practice factor, P, accounts for the effects of contour plowing, strip-cropping or terracing relative to straight-row farming up and down the slope. The cover-management factor, C, accounts for the specified crop and management relative to tilled continuous fallow. Several references on estimating these factors can be found online:

- USDA: RUSLE handbook (Renard et al., 1997)
- OMAFRA: USLE Fact Sheet http://www.omafra.gov.on.ca/english/engineer/facts/12-051.pdf
- U.N. Food and Agriculture Organization http://www.fao.org/docrep/T1765E/t1765e0c.htm

Watersheds / Subwatersheds

Watersheds outlets should correspond to reservoirs or other points of interest. This ensures that the sediment loads predicted by the model can be compared to observed data at these points. If known watershed maps exist, they should be used. Otherwise, watersheds and subwatersheds can be generated in ArcMap or QGIS based on the digital elevation model (see section on DEM for use of Fill tools to correct flow paths).

Exact locations of specific structures, such as reservoirs, should be obtained from the managing entity or may be obtained on the web at sites such as the National Inventory of Dams (http://geo.usace.army.mil/pgis/f?p=397:1:0). Global collections of dam locations and information include the Global Reservoir and Dam (GRanD) Database (http://www.gwsp.org/products/grand-database.html) and the World Water Development Report II dam database (http://wwdrii.sr.unh.edu/download.html.)

Calibration Parameters IC_0 and k_b

IC_0 and k_b are calibration parameters that define the relationship between the index of connectivity and the sediment delivery ratio (SDR). Vigiak et al. (2012) suggest that IC_0 is landscape independent and that the model is more sensitive to k_b . Advances in sediment modeling science should refine our understanding of the hydrologic connectivity and help improve this guidance. In the meantime, following other authors (Jamshidi et al., 2013), we recommend setting these parameters to their default values (IC_0 =0.5 and k_b =2), and using k_b only for calibration (Vigiak et al., 2012).

5.6.7 Appendix 2: Representation of Additional Sources and Sinks of Sediment

The InVEST model predicts the sediment deliver from sheetflow erosion, thus neglecting other sources and sinks of sediment (e.g. gully erosion, streambank, landslides, stream deposition, etc.), which can affect the valuation approach. Adding these elements to the sediment budget requires good knowledge of the sediment dynamics of the area and is typically beyond the scope of ecosystem services assessments. General formulations for instream deposition or gully formation are still an area of active research, with modelers systematically recognizing large uncertainties in process representation (Hughes and Prosser, 2003; Wilkinson et al., 2014). Consultation of the local literature to estimate the relative importance of additional sources and sinks is a more practical approach to assess their effect on the valuation approach.

	Process	Representation
Gully erosion (connected to the stream)	If the gully is connected to the stream, it is essentially a small stream tributary and its contribution may therefore be considered as streambank erosion.	The optional drainage layer can be used to represent this direct connection of the gully to the stream. To obtain the total sediment export, gully contribution (known from other studies) is added to the subwatershed level output.
Gully erosion (disconnected from the stream)	If the gully is disconnected, eroded soil is deposited on land and may eventually reach the stream together with rill-inter-rill sediment.	The additional soil loss can be added in post-processing. If a map of gullies and their estimated contribution is available (e.g. Vigiak 2011, Wilkinson 2014), it may be used with the SDR layer to compute the sediment export.
Streambank erosion	Bank erosion is a function of soil erodibility and shear stress, i.e. the hydraulic forces applied to the channel. Detailed knowledge the bank height, and soil resistance to friction are necessary to inform a simple process-based model of these process.	The additional soil loss can be added in post-processing. If local information is available on the magnitude of streambank erosion (e.g. per unit length of stream), users can compute this values for each subwatershed based on the stream layer generated by the model, and then add the value to the InVEST sediment export values.
Landslides	Landslides occur in areas of high rainfall intensity and instable slopes.	The additional soil loss can be added in post-processing.
Instream deposition	Larger particles tend to be deposited in the stream before reaching the catchment outlet	A simple method based on the Rouse number is proposed by Pelletier (2012), which relates deposition probability to the channel slope and settling velocity.

5.6.8 References

Bhattarai, R., Dutta, D., 2006. Estimation of Soil Erosion and Sediment Yield Using GIS at Catchment Scale. Water Resour. Manag. 21, 1635–1647.

Borselli, L., Cassi, P., Torri, D., 2008. Prolegomena to sediment and flow connectivity in the landscape: A GIS and field numerical assessment. Catena 75, 268–277.

Cavalli, M., Trevisani, S., Comiti, F., Marchi, L., 2013. Geomorphometric assessment of spatial sediment connectivity in small Alpine catchments. Geomorphology 188, 31–41.

Desmet, P.J.J., Govers, G., 1996. A GIs procedure for automatically calculating the USLE LS factor on topographically complex landscape units. J. Soi 51, 427–433.

De Vente J, Poesen J, Verstraeten G, Govers G, Vanmaercke M, Van Rompaey, A., Boix-Fayos C., 2013. Predicting soil erosion and sediment yield at regional scales: Where do we stand? Earth-Science Rev. 127 16–29

FAO, 2006. Guidelines for soil description - Fourth edition. Rome, Italy.

Hughes, A.O., Prosser, I.P., 2003. Gully and Riverbank erosion mapping for the Murray-Darling Basin. Canberra, ACT.

Jamshidi, R., Dragovich, D., Webb, A.A., 2013. Distributed empirical algorithms to estimate catchment scale sediment connectivity and yield in a subtropical region. Hydrol. Process.

Lopez-vicente, M., Poesen, J., Navas, A., Gaspar, L., 2013. Predicting runoff and sediment connectivity and soil erosion by water for different land use scenarios in the Spanish Pre-Pyrenees. Catena 102, 62–73.

Oliveira, A.H., Silva, M.A. da, Silva, M.L.N., Curi, N., Neto, G.K., Freitas, D.A.F. de, 2013. Development of Topographic Factor Modeling for Application in Soil Erosion Models, in: Intechopen (Ed.), Soil Processes and Current

Trends in Quality Assessment. p. 28.

Pelletier, J.D., 2012. A spatially distributed model for the long-term suspended sediment discharge and delivery ratio of drainage basins 117, 1–15.

Renard, K., Foster, G., Weesies, G., McCool, D., Yoder, D., 1997. Predicting Soil Erosion by Water: A Guide to Conservation Planning with the revised soil loss equation.

Renard, K., Freimund, J., 1994. Using monthly precipitation data to estimate the R-factor in the revised USLE. J. Hydrol. 157, 287–306. Roose, 1996. Land husbandry - Components and strategy. Soils bulletin 70. Rome, Italy.

Sougnez, N., Wesemael, B. Van, Vanacker, V., 2011. Low erosion rates measured for steep , sparsely vegetated catchments in southeast Spain. Catena 84, 1–11.

Tarboton, D., 1997. A new method for the determination of flow directions and upslop areas in grid digital elevation models. Water Resour. Res. 33, 309–319.

Vigiak, O., Borselli, L., Newham, L.T.H., Mcinnes, J., Roberts, A.M., 2012. Comparison of conceptual landscape metrics to define hillslope-scale sediment delivery ratio. Geomorphology 138, 74–88.

Wilkinson, S.N., Dougall, C., Kinsey-Henderson, A.E., Searle, R.D., Ellis, R.J., Bartley, R., 2014. Development of a time-stepping sediment budget model for assessing land use impacts in large river basins. Sci. Total Environ. 468-469, 1210–24.

5.7 Unobstructed Views: Scenic Quality Provision

5.7.1 Summary

The natural and scenic views of marine and coastal seascapes can contribute to the well-being of local communities in a number of ways. Scenic amenities play an important role in augmenting local economies by attracting visitors who support local businesses. The value of local property partially depends on attributes of its location and scenic views often increase local property values (Sanders and Polasky 2009, Bourassa et al. 2004, Benson et al. 2004). Local communities and their residents often become strongly attached to views and show fervent opposition to new development that has the potential to threaten the integrity of existing views and diminish the benefits drawn from those views (Ladenburg and Dubgaard 2009, Haggett 2011). The InVEST scenic quality model allows users to determine the locations from which new nearshore or offshore features can be seen. It generates viewshed maps that can be used to identify the visual footprint of new offshore development. Inputs to the viewshed model include: topography and bathymetry, locations of offshore facilities of interest, and the locations of viewers (e.g. population centers or areas of interest such as parks or trails). The model does not quantify economic impacts of altering the viewshed, but it can be adapted to compute viewshed metrics for use in a more detailed valuation study. A key limitation of the model is that it does not currently account for the ways in which vegetation or land-based infrastructure may constrain land areas that are visually affected by offshore development.

5.7.2 Scenic Quality Standalone Beta

Currently we are working on the next generation platform of InVEST (3.0) and deploying parts of it as prototype InVEST models. You can try out the 3.0 version of Scenic Quality by navigating to your Windows Start Menu -> All Programs -> InVEST +VERSION+ -> Scenic Quality. The interface does not require ArcGIS and the results can be explored with any GIS tool including ArcGIS, QuantumGIS, and others.

In an earlier version of InVEST this tool didn't measure the economic impact of visual disamenities (i.e. valuation). The new 3.0 version performs valuation, and the user can notice the new *valuation* tab for the new valuation parameters besides the *general* tab that regroups all the parameters in the previous version of the model.

5.7.3 Introduction

Coastal ecosystems are increasingly dominated by human activities. This rise in human activities can compromise the unique scenic qualities associated with coastal and marine areas. The coastline and 'seascape' is an important economic asset that attracts visitors for tourism and recreation and contributes to the general quality of life for people living near the coast. Near and offshore development projects often raise considerable concern within the local communities that value the natural seascape for its inherent beauty. Visual impacts are external effects that unless measured and accounted for, do not factor into the calculus of weighing the costs and benefits of new coastal development. Applications using viewshed analysis range from the siting of aquaculture facilities to minimize spatial competition with tourism activities (Perez 2003) to seascape and shoreline visibility assessment of offshore wind projects (Environmental Design and Research 2006). Because scenic beauty is an attribute generally considered to be important to people living near the coast and for those who visit coastal areas to enjoy the ocean and the marine environment, coastal planners can incorporate measures of visual amenities and/or disamenities into broader policy deliberations and planning exercises. Because most applications of viewshed analysis involve examining the negative impacts of new facilities, language within the InVEST scenic quality model assumes the objects viewed have a negative impact on views. However, positive interpretation of viewing these objects can be included with interpretation of model results.

The InVEST scenic quality model provides users with a simple way to provide information about potential tradeoffs between nearshore and offshore development proposals and the visual impacts of those projects. The viewshed maps produced by the model can be used to identify coastal areas that are most likely to be directly affected by additions to the seascape. They can serve as valuable input into broader analyses that consider a range of services provided by the marine environment.

This model can be used to compute the costs associated with offshore visual impacts, these costs are likely to decrease as the location of facilities moves further offshore, while the costs of installing and operating offshore facilities generally increase with distance from the shoreline. The few valuation studies that explore the economic magnitude of visual disamenities resulting from offshore development projects show a complex picture. One recent study found that individuals living along the coast have external costs ranging from \$27 to \$80 resulting from the visual disamenity of an offshore wind project (Krueger et al. 2010). In contrast, Firestone et al. (2009) found that public acceptance for offshore renewable energy projects is growing and may be less contentious than previously anticipated.

5.7.4 The Model

The scenic quality model provides information about the visibility of offshore objects from the surrounding landscape or seascape. Offshore and nearshore development projects, such as renewable wave energy facilities or aquaculture facilities, have the potential to impact the visual amenities that are an important feature of many coastal areas. The results of viewshed analysis will be useful for decision-makers who would like to identify areas where visual impacts may be an important factor to incorporate into planning.

The model inputs are groupped in two tabs: the **General** tab has all the entries necessary to run the viewshed computation such as the location of a DEM and a point shapefile that identifies the locations of sites that contribute to visual impacts. The **Valuation** tab allows the user to select the functional form of the valuation function, and its parameters. The viewshed analysis is then computed over a user-defined area of interest (AOI).

The model will create as many as four outputs that can be used to assess the visible impact of any type of facility added to the marine environment. The first output, "vshed", is a valuation raster based on the visual quality at any given pixel. If the valuation is set to the constant 1 independent of the viewing distance, the raster reduces to a record of the number of sites (e.g. wave energy facilities or aquaculture farms) that are visible from a given raster cell on the land or seascape. The cells are then classified using quantiles to produce "vshed_qual" with the following class breaks: 1. Unaffected, 2. Low Visual Impact/High Visual Quality, 3. Moderate Visual Impact/Medium Visual Quality, 4. High Visual Impact/Low Visual Quality, 5. Very High Visual Impact/Poor Visual Quality.

The third output computes the resident population that falls within the viewshed of any facility. The model uses the Global Rural-Urban Mapping Project (GRUMP) gridded population of the world data (CIESIN 2004) to compute the number of residents who are unaffected by the facility (or facilities) and the number of residents who live in areas that fall within the viewshed of at least one facility. The population counts are tabulated in the "populationStats.html" file found in the output folder. Users should note that this globally available population data does not account for seasonal or daily users in an area. Alternatively, you can provide your own population raster data.

The final optional output allows for the examination of the visual impacts on areas of interest where the view is of particular concern (e.g. parks, trails, marine reserves). It utilizes a user-defined set of polygons and computes the percent area within each polygon from which at least one offshore site is visible. Each polygon is then classified by the percentage of that polygon's area that is visually impacted by offshore developments. These results can be used to identify and rank areas according to visual impacts.

Additional files are created for each feature *X* at each step of the computation: - *visibility_X.tif* is the visibile area raster computed from feature X. - *distance_X.tif* is the distance in meters from feature X across the visible area in *visibility_X.tif*. - *viewshed_X.tif* is computed from *distance_X.tif* where the valuation function is applied to the distances. - *vshed_X.tif* is *viewshed_X.tif* scaled by the coefficient associated to feature X. If no coefficient is specified, the model assumes a coefficient of 1.0, in which case *vshed_X.tif* is identical to *viewshed_X.tif*. The file *vshed.tif* is the sum of each individual *sched_X.tif*.

How it Works

The InVEST scenic quality model computes the visual impact of features in the landscape in four steps: 1. Visibility calculation: the model compute a visibility raster for each point feature *X* (*visibility_X.tif*). It implements a simple line of sight algorithm that only computes visibility along the lines originating from the viewpoint to the center of the perimeter raster cells. 2. Valuation: The model applies a valuation function (either logarithmic or third order polynomial) across the visibility raster using the distance to the point feature (*viewshed_X.tif*). 3. Weighting: Each feature in the point shapefile can have a field *coeff*, which is used to scale the values returned by the valuation function (*vshed_X.tif*). 4. Summation: The weighted rasters are summed up to produce the visual impact output raster *vshed.tif*.

The valuation function is either logarithmic:

$$f(x) = a + b \cdot log(x) \tag{5.27}$$

or a third degree polynomial:

$$f(x) = a + b \cdot x + c \cdot x^2 + d \cdot x^3 \tag{5.28}$$

Where x is the distance from the cell center to a point feature, and a, b, c, and d are coefficients. With the default parameter values (*a=1*, *b=c=d=0*), the model computes an aggregate viewshed. The valuation function is computed up to a maximum valuation radius that defaults to 8000 meters. For short distnaces, the logarithmic and polynomial forms could degenerate to unrealistic high values. To avoid this situation, the model uses a linear function l(x):

$$l(x) = A \cdot x + B \tag{5.29}$$

where A = f'(1000) and B = f(1000). Since the function quantifies dollar amounts, it should be positive throughout its range. The model will test if the function is positive at the maximum radius, and return an error if not.

5.7.5 Limitations and Simplifications

The global DEM included with the scenic quality model does not account for trees, buildings, or other structures that can obscure the view. If users have a raster layer that represents the locations of trees, buildings, or other obstructions (and their heights) this information can be incorporated into the DEM to create a more realistic surface to obscure or allow views. The model does account for the curvature of the earth in limiting the line of sight but it does not limit the distance at which objects of varying size and quality may be visible to the human eye in the default settings. As long as there is a straight-line vector that can be computed from a particular DEM grid cells to any offshore point, that grid cell will be counted as visible. This should be carefully considered when interpreting viewshed impact maps from facilities located far offshore when default settings are used. However, users can provide an outer radius that limits the search distance when identifying areas visible from each offshore development site.

5.7.6 Data Needs

The model's interface is composed of two tabs, **General** and **valuation**. The former contains all the inputs necessary to run the basic viewshed analysis (without valuation), as well as additional inputs for population data and specific impacted areas. The latter tab contains the valuation-related inputs. Here we describe each option in more detail.

General Tab

1. **Workspace (required).** Users are required to specify a workspace folder path. It is recommend that the user create a new folder for each run of the model. For example, by creating a folder called "runBC" within the "ScenicQuality" folder, the model will create "intermediate" and "output" folders within this "runBC" workspace. The "intermediate" folder will compartmentalize data from intermediate processes. The model's final outputs will be stored in the "output" folder.

```
Name: Path to a workspace folder. Avoid spaces.
Sample path: \InVEST\ScenicQuality\runBC
```

2. **Area of Interest (AOI) (required).** An AOI instructs the model where to clip the input data and the extent of analysis. Users will create a polygon feature layer that defines their area of interest. The AOI must intersect the Digital Elevation Model (DEM).

```
File type: polygon shapefile (.shp)
Sample path: \InVEST\ScenicQuality\AOI_WCVI.shp
```

3. **Cell Size (meters) (optional).** This determines the spatial resolution at which the model runs and at which the results are summarized. For example, if you want to run the model and see results at a 100m x 100m grid cell scale then enter "100." You can only define a resolution that is equal to or coarser than the model's native resolution as established by the current DEM (input # 4). If you want to run the model and produce output at the current DEM's resolution (the model's native resolution) you can leave this input field blank. The coarser the scale (and larger the number), the faster the model runs.

```
Type: text string (direct input to the interface)
Sample (default): 500
```

4. **Features Impacting Scenic Quality (required).** The user must specify a point feature layer that indicates locations of objects that contribute to negative scenic quality, such as aquaculture netpens or wave energy facilities. Users wish to including polygons (e.g. clear-cuts) in their analysis must convert the polygons to a grid of evenly spaced points.

```
File type: point shapefile (.shp)
Sample path: \InVEST\ScenicQuality\AquaWEM_points.shp
```

The model will compute a viewshed for each feature separately and aggregate them into a combined viewshed. The user can specify up to three fields (all fields are optional) to assign a maximum viewing distance, a viewshed importance coefficient, or a viewpoint height to each feature:

- *Maximum viewing distance*: Integer field named either "RADIUS" (preferred, case insensitive) or "RADIUS2" (kept for backwards compatibility) specifying the maximum length of the line of sight in meters originating from a viewpoint. The value can either be positive (preferred) or negative (kept for backwards compatibility), but is converted to a positive number. The model assumes a value of 8000m (8km) if the field doesn't exist.
- *Viewshed importance coefficient*: The user can assign an importance to each viewshed by scaling them with a real number (either positive or negative) stored in the field "coeff". The model assumes a coefficient of 1.0 if the field doesn't exist.
- *Viewpoint height*: Each feature elevation above the ground can be specified as a positive real number in the field "height". The default value is 0.0 if the field doesn't exist.

5. **Digital Elevation Model (DEM) (required).** A global raster layer is required to conduct viewshed analysis. Elevation data allows the model to determine areas within the AOI's land-seascape where features from input #4 are visible.

```
Format: standard GIS raster file (e.g., ESRI GRID or IMG), with elevation values
Sample data set: \InVEST\ScenicQuality\Base_Data\Marine\DEMs\claybark_dem
```

6. **Refractivity Coefficient (required).** The earth curvature correction option corrects for the curvature of the earth and refraction of visible light in air. Changes in air density curve the light downward causing an observer to see further and the earth to appear less curved. While the magnitude of this effect varies with atmospheric conditions, a standard rule of thumb is that refraction of visible light reduces the apparent curvature of the earth by one-seventh. By default, this model corrects for the curvature of the earth and sets the refractivity coefficient to 0.13.

```
Format: A string of numeric text with a value between 0 and 1
Sample (default): 0.13
```

7. **Population Raster (required).** A global raster layer is required to determine population within the AOI's land-seascape where features from input #4 are visible and not visible.

```
Format: standard GIS raster file (ESRI GRID) with population values
Sample data set (default): \InVEST\Base_Data\Marine\Population\global_pop
```

8. **Overlap Analysis Features (optional).** The user has the option of providing a polygon feature layer where they would like to determine the impact of points (input #4) on visual quality. This input must be a polygon and projected in meters. The model will use this layer to determine what percent of the total area of each feature can see at least one of the points from input #4.

```
File type: polygon shapefile (.shp)
Sample path: \InVEST\ScenicQuality\BC_parks.shp
```

Valuation Tab

9. **Valuation function.** Type of economic function the user wishes to use to quantify the visual impact of disamenities. The coefficients for each function can be specified in the following inputs.

```
Format: An item selected from a drop-down menu
Default: Polynomial
```

10. **'a' coefficient.** Constant value (independent of the distance *x*) used by both the logarithmic and the polynomial functions. It is set to 1.0 by default.

```
Format: A string of numeric text (direct input to the interface)
Default: 1.0
```

11. **'b' coefficient.** Coefficient used by both the logarithmic and the polynomial form. It weights the first order factor in the polynomial form, and the logarithmic factor for the logarithmic function. It is set to 0.0 by default.

```
Format: A string of numeric text (direct input to the interface)
Default: 0.0
```

12. **'c' coefficient.** Coefficient used in the polynomial form only to weight the second order term. It has no effect if the user chooses the logarithmic valuation function. It is set to 0.0 by default.

```
Format: A string of numeric text (direct input to the interface)
Default: 0.0
```

13. **'d' coefficient.** Coefficient that weights the third order factor. It is set to 0.0 by default.

```
Format: A string of numeric text (direct input to the interface)
Default: 0.0
```

14. **Maximum valuation radius.** Valuation will only be computed for cells that fall within the maximum valuation radius. The maximum radius is a positive number in meters.

```
Format: A string of numeric text (direct input to the interface)
Default: 8000.0
```

5.7.7 Running the Model

The model is available as a standalone application accessible from the Windows start menu. For Windows 7 or earlier, this can be found under *All Programs -> InVEST +VERSION+ -> Scenic Quality*. The standalone can also be found directly in the InVEST install directory under the subdirectory *invest-3_x86/invest_scenic_quality.exe*.

Viewing Output from the Model

Upon successful completion of the model, a file explorer window will open to the output workspace specified in the model run. This directory contains an *output* folder holding files generated by this model. Those files can be viewed in any GIS tool such as ArcGIS, or QGIS. These files are described below in Section *Interpreting Results*.

Final Results

Output Folder

- Output\vshed_qual
  - This raster layer contains a field that classifies based on quartiles the visual quality within the AOI. The visual quality classes include: unaffected (no visual impact), high (low visual impact), medium (moderate visual impact), low (high visual impact), and very low (very high visual impact).
  - Additionally, the range of sites visible for each visual quality class is specified in this output's attribute table.
  - This layer can be symbolized by importing the symbology from the file "\ScenicQuality\Input\vshed_qual.lyr"
- Output\vshed
  - This raster layer is the original output after the viewshed tool is run. It contains values ranging from 0 to the total number of points visible from each cell on the land or seascape. For example, all cells with a value of "4" would indicate that at that location four points are visible.
  - In order to compare scenario runs, use this layer rather than vshed_qual. By calculating the difference between "vshed" outputs from multiple runs, a user can assess changes in visual quality across scenarios.
- Output\vp_overlap.shp
  - This polygon feature layer contains a field called "AreaVShed" which expresses the percentage of area within each polygon where at least one point contributing to negative scenic quality is visible as compared to the total area of that polygon.
  - This layer can easily be symbolized by importing the symbology from the file "\ScenicQuality\Input\vp_overlap.lyr"
- Output\populationStats.html

  - This html file includes a table and indicates the approximate number of people within the AOI that are 1) unaffected (no sites contributing to negative scenic quality are visible) and 2) affected (one or more sites visible).
- scenic_quality_log-[yr-mon-day-hour-min-sec].txt
  - Each time the model is run a text file will appear in the workspace folder. The file will list the parameter values for that run and be named according to the date and time.

Intermediate Folder

- intermediate\aoi_dem
  - Reprojected AOI to match the DEM's projection. Used for clipping the DEM to the AOI.
- intermediate\dem_vs
  - DEM clipped with the AOI using aoi_dem.
- intermediate\dem_vs_re
  - Raster similar to DEM, with negative heights set to zero.
- intermediate\vshed_bool
  - Raster where only the pixels spanned by the viewsheds are set to 1, and the others are 0.
- intermediate\aoi_pop
  - Reprojected AOI to match the population layer's projection. Used for clipping the population layer to the AOI.
- intermediate\pop_clip
  - Population raster clipped to the AOI.
- intermediate\pop_prj
  - Clipped population raster reprojected to the original population layer.
- intermediate\pop_vs
  - Population raster resampled to the user-defined cell size and aligned to the AOI.

5.7.8 References

Benson E., Hansen, J., Schwartz, A., and Smersh, G., 1998. Pricing residential amenities: the value of a view. Journal of Real Estate Research, 16: 55-73.

Bourassa, S., Hoesli, M. and Sun, J. 2004. What's in a view? Environment and Planning A. 36(8): 1427-1450.

Center for International Earth Science Information Network (CIESIN), Columbia University; International Food Policy Research Institute (IFPRI); The World Bank; and Centro Internacional de Agricultura Tropical (CIAT). 2004. Global Rural-Urban Mapping Project (GRUMP), Alpha Version: Population Grids. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University. Available at http://sedac.ciesin.columbia.edu/gpw. (downloaded on 1/6/2011).

Environmental Design and Research, P.C. 2006. Seascape and shoreline visibility assessment. Cape Wind Energy Project. Cape Cod, Martha's Vineyard, and Nantucket, Massachusetts. Prepared for Cape Wind Associates, L.L.C. Boston, Mass. Syracuse, N.Y. July 2006.

Firestone, J., Kempton, W. & Krueger, A., 2009. Public acceptance of offshore wind power projects in the USA. Wind Energy, 12(2):183-202.

Haggett, C. 2011. Understanding public responses to offshore wind power. Energy Policy. 39: 503-510.

Krueger, A., Parson, G., and Firestone, J., 2010. Valuing the visual disamenity of offshore wind power at varying distances from the shore: An application of on the Delaware shoreline. Working paper. Available at: http://works.bepress.com/george_parsons/doctype.html.

Ladenburg, J. & Dubgaard, A., 2009. Preferences of coastal zone user groups regarding the siting of offshore wind farms. Ocean & Coastal Management, 52(5): 233-242.

Perez, O.M., Telfer, T.C. & Ross, L.G., 2005. Geographical information systems-based models for offshore floating marine fish cage aquaculture site selection in Tenerife, Canary Islands. Aquaculture Research, 36(10):946-961.

Sander, H.A. & Polasky, S., 2009. The value of views and open space: Estimates from a hedonic pricing model for Ramsey County, Minnesota, USA. Land Use Policy, 26(3):837-845.

5.8 Visitation: Recreation and Tourism

5.8.1 Summary

Recreation and tourism are important components of many national and local economies and they contribute in innumerable ways to quality of life, sense of place, social connection, physical wellbeing, learning, and other intangibles. To quantify the value of natural environments, the InVEST recreation model predicts the spread of person-days of recreation, based on the locations of natural habitats and other features that factor into people's decisions about where to recreate. The tool estimates the contribution of each attribute to visitation rate in a simple linear regression. In the absence of empirical data on visitation, we parameterize the model using a proxy for visitation: geotagged photographs posted to the website flickr. Using photo-user-day estimates, the model predicts how future changes to natural features will alter visitation rates. The tool outputs maps showing current patterns of recreational use and maps of future patterns of use under alternate scenarios.

5.8.2 Introduction

Recreation and tourism are important components of many national and local economies and they contribute in innumerable ways to quality of life, sense of place, social connection, physical wellbeing, learning, and other intangibles. A key reason for studying patterns of recreation or tourism is the economic significance of this industry. The total contribution of travel and tourism to the world's gross domestic product (GDP) in 2011 was approximately $6 B USD (9% of GDP), with expected growth to $10 B USD by 2022 (World Travel and Tourism Council 2012). Of course, economic impacts are only one way of measuring the importance of recreation and tourism. These activities are critical contributors to diverse aspects of human wellbeing (Russell et al. 2013). For example, outdoor recreation is a spiritual experience for many people (eg Trainor and Norgaard 1999), and social interactions in nature contribute to building a sense of place (e.g. Willox et al. 2012).

A major and growing portion of recreation is "nature-based", involving interactions with or appreciation of the natural environment (Balmford et al. 2009). For these types of activities, characteristics of the environment influence people's decisions about where, when, and how to recreate. SCUBA divers, for example, select destinations based on the water clarity, water temperature, and diversity of marine life (Williams and Polunin 2000, Uyarra et al. 2009). Birdwatchers are drawn to the best places to see target species (Naidoo and Adamowicz 2005), which inevitably are places where natural systems support populations of desirable birds (Puhakka et al. 2011). Some recreation depends on environmental attributes such as species richness (Loureiro et al. 2012), the diversity of habitats (Neuvonen et al. 2010, Loureiro et al. 2012), precipitation (Loomis and Richardson 2006), and temperature (Richardson and Loomis 2005), as well as to other attributes such as infrastructure and cultural attractions (Mills and Westover 1987, Hill and Courtney 2006).

5.8.3 The Model

The purpose of the InVEST recreation model is to predict the spread of person-days of recreation, based on the locations of natural habitats, accessibility, and built features such as roads that factor into people's decisions about where to recreate (Adamowicz et al. 2011). The tool outputs maps showing current patterns of recreational use and, optionally, maps of future use under alternative scenarios.

How it Works

The tool estimates the contribution of attributes of the landscape to the visitation rate, using a simple linear regression:

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} \text{ for } i = 1...n,$$

where x_{ip} is the coverage of each attribute in each cell, i, within an Area of Interest (AOI) containing n cells. In the absence of empirical data on visitation for y_i, we parameterize the model using a proxy for visitation: geotagged photographs posted to the website flickr (see *Photograph User Days* section for more information). Stated again, the InVEST recreation model predicts the spread of person-days of recreation in space. It does this using attributes of places, such as natural features (eg parks), built features (eg roads), and human uses (eg industrial activities), among others.

The tool begins by log-transforming all y_i values, by taking the natural log of average photo-user-days per cell + 1. Then, a simple linear regression is performed to estimate the effect of each attribute on log-transformed visitation rates across all grid cells within the study region. These estimates (the β_p values) can be used for additional model runs, via the InVEST recreation scenario tool, to predict how future changes to the landscape will alter visitation rate. The model uses ordinary least squares regression, performed by the lm function for R (R Core Team 2013).

Photograph User Days

Since fine-scale data on numbers of visitors is often only collected at only a few specific locations in any study region, we assume that current visitation can be approximated by the total number of annual person-days of photographs uploaded to the photo-sharing website flickr. Many of the photographs in flickr have been assigned to a specific latitude/longitude. Using this location, along with the photographer's user name and date that the image was taken, the InVEST tool computes the total annual days that a user took at least one photograph within each cell, then returns to users the average annual number of photo-user-days from 2005-2012. We have observed that the number of recreators who visit a location annually is related to the number of photographs taken in the same area and uploaded to the flickr database at 836 visitor attractions worldwide (Wood et al. 2013). The density of photographs varies spatially, and this has ramifications for the cell-size that can be chosen for analysis (see *Initial Tool*: Cell size).

Predictor Variables

We find that it often helps to consider at least one variable from several main categories: natural capital (eg habitats, lakes), built capital (eg roads, hotels), industrial activities, and access or cost (eg distance to major airport). Often, single variables representing each of these categories can explain the majority of variation in photo-user-days. To facilitate this, the tool comes pre-loaded with several optional sources of global spatial data including total population and natural habitats on land and in the ocean (described in the *Data Needs* section). The tool also allows users to upload their own spatial data (in any vector shapefile format), if they have information on additional or alternative attributes that might be correlated to people's decisions about where to recreate.

Limitations and Simplifications

The model does not presuppose that any predictor variable has an effect on visitation. Instead, the tool estimates the magnitude of each predictor's effect based on its spatial correspondence with current visitation in the area of interest.

The values of photo-person-days per cells are taken as a proxy-measure of visitation and are regressed against the values of the predictor variables across all cells. In subsequent model-runs, the tool employs the beta values computed in the initial model-run to predict visitation, given a spatial configuration of the predictors, under future scenarios. This step requires the assumption that people's responses to attributes that serve as predictors in the model will not change over time. In other words, in the future, people will continue to be drawn to or repelled by the attributes as they are currently.

5.8.4 Data Needs

The following outlines the options presented to the user via the two interfaces, and the content and format of the required and optional input data used by the model. More information on how to format and obtain data is provided in *Appendix A*.

Initial Tool

1. **Workspace (required).** Users must specify a path to the workspace folder where the tool will create a file of results:

```
Name: Path to a workspace folder.  Avoid spaces.
Sample path: \InVEST\Recreation\
```

2. **Area of Interest (required).** This input provides the model with a geographic shape of the area of interest (AOI). The AOI must be projected (see *Supported Projections*) and have an associated linear unit. The extent of the AOI is used to create the grid (if checked, see below) and only cells that fall within the AOI are included. The total area of the AOI must be smaller than 800,000 square km:

```
Name: File can be named anything, but no spaces in the name
File type: polygon shapefile (.shp)
```

3. **Grid type (required).** This input specifies the shape of the grid cells. Rectangular grids contain squares oriented parallel to the coordinate system of the AOI. Hexagonal grids contain hexagons oriented with a long diagonal parallel to the horizontal component of the coordinate system.

4. **Cell size (required).** This input specifies the size of grid cells. The cell size is **in the same linear units as the AOI**. For example, if the AOI is in a UTM projection with units of meters, and cell size parameter will also be in meters. The minimum allowable grid cell size is three square km and the AOI must contain at least five cells.

   The appropriate size and number of cells depends on several factors, including the goals of the study and the density of photographs, which varies from region to region. In order for the model to compute the effects of predictor variables (as described in the *How it Works* section), users must select a sufficiently large cell size, such that the majority of cells contain photographs. We recommend that users begin by running the model with cells ranging between 100-1000 square km. Then, iteratively assess the model outputs (grid.shp and regression_summary.pdf, described in *Interpreting Results*) and re-run the model to determine an appropriate cell size.

5. **Comments (optional).** This input provides the model with text comments to include with the outputs.

6. **Data Directory (optional).** Users can optionally specify a data folder containing additional geographic data to use as predictors (for x_{ip} values described in *How it Works*). The data can be in a geographic or projected coordinate system, but it must be known and specified in the projection file (.prj). Additionally, the geographic data can be classified if an optional classification table (.csv) is specified (see *Categorization Tables* for more information):

```
Name: Path to a data directory.  Avoid spaces.
Sample path: \InVEST\Recreation\data\BC\pred
```

7. **Download Data (optional).** User can choose have the processed predictors, including the user supplied predictors, returned with the model results.

8. **Global Default Data (optional).** The tool provides several global spatial datasets which users can optionally include as predictor variables for their AOI. Further information on these datasets is available in the *Default Predictors* Section of Appendix A.

   - **2010 Population (optional).** Oak Ridge National Laboratory LandScan (2010) population data. Please note that due to the license agreement, these data cannot be included in downloaded data.
   - **OSM Points (optional).** Open Street Map (2012) point features categorized into cultural, industrial, natural, structural, and miscellaneous features. See *OSM Categorization*.
   - **OSM Lines (optional).** Open Street Map (2012) line features categorized into cultural, industrial, natural, structural, and miscellaneous features. See *OSM Categorization*.
   - **OSM Polygons (optional).** Open Street Map (2012) polygon features categorized into cultural, industrial, natural, structural, and miscellaneous features. See *OSM Categorization*.
   - **Protected Areas (optional).** UNEP-WCMC World Data Base on Protected Areas (2012) polygon features.
   - **LULC (optional).** ESA GlobCover (2008) land use and land cover data. See LULC categorization.
   - **Mangroves (optional).** UNEP-WCMC Ocean Data Viewer Mangroves (1997).
   - **Coral Reefs (optional).** UNEP-WCMC Ocean Data Viewer Coral Reefs (2010).
   - **Seagrasses (optional).** UNEP-WCMC Ocean Data Viewer Seagrasses (2005).

Scenario Tool

1. **Workspace (required).** Users must specify a path to the workspace folder. The model will create a file of results here:

```
Name: Path to a workspace folder.  Avoid spaces.
Sample path: \InVEST\Recreation\
```

2. **init.json (required).** The configuration file created by the Initial Tool and saved in the results folder in the initial workspace.

3. **Data Directory (required).** Users must specify a data folder that contains the modified predictors for the scenario. Uploaded shapefiles must have identical names as those uploaded for the first run using the Initial Tool. It is only necessary to provide the changed shapefiles for scenario runs, unchanged data can be read from the initial model run. The data can be in a geographic or projected coordinate system, but it must be known and specified in the projection file (.prj). Additionally, the geographic data can be classified if an optional classification table (.csv) is specified (see the *Categorization Tables* Section for more information):

```
Name: Path to a data directory.  Avoid spaces.
Sample path: \InVEST\Recreation\data\BC\pred
```

4. **Comments (optional).** This input provides the model with text comments to include with the outputs.

5.8.5 Running the Model

Warning: The recreation model requires a connection to the internet.

The model uses an interface to input all required and optional data (see *Data Needs*), which are then sent to a server managed by the Natural Capital Project in California, where computations are performed. Consequently, this model requires a connection to the internet. The server outputs a vector polygon shapefile and .csv tables of results (described in *Interpreting Results*). The InVEST recreation model consists of two individual tools, which must be run consecutively:

1. The Initial tool, which computes photo-user-days (y_i), coverages of predictors (x_{ip}), and effects of predictors (β_p).
2. The Scenario tool, which uses effects per predictor (β_p) to estimate future visitation rates.

The time required to run the Initial Tool varies depending on the extent of the AOI, the number grid cells, and the number and resolution of predictor layers. The Scenario Tool takes less time to run.

Please note, the server performing the analysis also records the IP address of each user.

5.8.6 Interpreting Results

Model Outputs

The follwing is a short decription of each of the outputs from the Scenario model. Each of these output files is saved in the outputs saved into the workspace directory in a file named *results-YYYY-MM-DD–HH_MM_SS.zip* where *YYYY-MM-DD–HH_MM_SS* represents the year, month, day, hour, minute, and seconds, respectively.

- aoi_params.csv
  - This text file contains the parameters estimated by the linear regression (see *How it Works*), including the β_p and p values. Each predictor variable must be present in cells within the AOI in order to estimate their effects. Any predictor variables that cannot be estimated remain blank in the aoi_params.csv table.
- comments.txt
  - This text file contains the optional user comments.
- grid.shp
  - This polygon feature layer contains the gridded AOI with the number of photo-user-days and coverage of each predictor variable per cell.
  - USDYAV is the average photo-user-days per year (using all photos from 2005-2012). This corresponds to the average *PUD* described in Wood et al. (2013).
  - USDYAV_PR is simply the proportion of total USDYAV per cell.
  - USDYAV_EST is the average photo-user-days estimated by the linear regression equation.
- init.json
  - This configuration file contains the initial tool parameters. It should not be edited.
- download/ (optional)
  - This folder contains the feature layers for processed predictors.

5.8.7 Appendix A

Supported Projections

The supported projections are a subset of the European Petroleum Survey Group (EPSG) projections, which are commonly used and supported across a wide range of industries and platforms. Specifically we support the EPSG projections that use linear units (meters, feet, etc.) also known as projected coordinate systems, which include the following:

- Universal Transverse Mercator projections
- Albers projections
- Lambert projections

and many more.

For more information on EPSG projections see http://spatialreference.org/ref/epsg/.

Depending on the source of the data there can be minor variations in how a projection is stored, which may raise a projection error. If you have a projection that uses linear units and it is not working with the recreation model, please start a discussion on the user forum at http://ncp-yamato.stanford.edu/natcapforums/.

Predictor Variables

Upload Directory

Predictor folders should contain *predictors for the model run only*. Files must be ESRI shapefiles format. All files must be under 20MB zipped and file names are limited to US-ASCII and cannot contain accent marks. Finally, the following file names are reserved for internal use and cannot be used: *borders*, *duplicates*, *photos*, *planet_osm*, *predictor*, *prj*, *searches*, *spatial*, *srid*, *tmp*, *users*, *wkt*.

Categorization Tables

Categorization Tables are tab delmited text files with three required columns: the field name, the field value, and the category name. The table should contain a row header and the category names cannot contain spaces or symbols.

OSM Categorization

A supplementary table provides the categorization scheme used for all OSM features. It is not exhaustive, but almost all other features fall into another cateogry. For more information on how OSM features are tagged see the OSM wiki.

LULC Classification

The following is the reclassification table used for the global land use and land cover.

Value	Re-classify	Label	Red	Green	Blue
11	agri-culture	Post-flooding or irrigated croplands (or aquatic)	170	240	240
14	agri-culture	Rainfed croplands	255	255	100
20	agri-culture	Mosaic cropland (50-70%) / vegetation (grassland/shrubland/forest) (20-50%)	220	240	100
30	agri-culture	Mosaic vegetation (grassland/shrubland/forest) (50-70%) / cropland (20-50%)	205	205	102
40	forest	Closed to open (>15%) broadleaved evergreen or semi-deciduous forest (>5m)	0	100	0
50	forest	Closed (>40%) broadleaved deciduous forest (>5m)	0	160	0
60	forest	Open (15-40%) broadleaved deciduous forest/woodland (>5m)	170	200	0
70	forest	Closed (>40%) needleleaved evergreen forest (>5m)	0	60	0
90	forest	Open (15-40%) needleleaved deciduous or evergreen forest (>5m)	40	100	0
100	forest	Closed to open (>15%) mixed broadleaved and needleleaved forest (>5m)	120	130	0
110	shrub-land	Mosaic forest or shrubland (50-70%) / grassland (20-50%)	140	160	0
120	grass-land	Mosaic grassland (50-70%) / forest or shrubland (20-50%)	190	150	0
130	shrub-land	Closed to open (>15%) (broadleaved or needleleaved, evergreen or deciduous) shrubland (<5m)	150	100	0
140	grass-land	Closed to open (>15%) herbaceous vegetation (grassland, savannas or lichens/mosses)	255	180	50
150	bare	Sparse (<15%) vegetation	255	235	175
160	forest	Closed to open (>15%) broadleaved forest regularly flooded (semi-permanently or temporarily) - Fresh or brackish water	0	120	90
170	water	Closed (>40%) broadleaved forest or shrubland permanently flooded - Saline or brackish water	0	150	120
180	water	Closed to open (>15%) grassland or woody vegetation on regularly flooded or waterlogged soil - Fresh, brackish or saline water	0	220	130
190	urban	Artificial surfaces and associated areas (Urban areas >50%)	195	20	0
200	bare	Bare areas	255	245	215
210	water	Water bodies	0	70	200
220	snow and ice	Permanent snow and ice	255	255	255
230	no data	No data (burnt areas, clouds,…)	0	0	0

Default Predictors

The default global predictor data provided by the Initial and Scenario Tools are from the following sources.

Predictor	Description	Year	Source
2010 Population	A global distribution of ambient population (average over 24 hours).	2010	http://www.ornl.gov/sci/landscan/
OSM Points	Open Street Map point features.	2012	http://www.openstreetmap.org/
OSM Lines	Open Street Map line features.	2012	http://www.openstreetmap.org/
OSM Polygons	Open Street Map polygon features.	2012	http://www.openstreetmap.org/
LULC	Globcover v 2.2 global land cover map for the period from December 2004 to June 2006.	2008	http://due.esrin.esa.int/globcover/

5.8.8 References

Adamowicz, WL, R Naidoo, E Nelson, S Polasky, J Zhang. 2011. Nature-based tourism and recreation. In: Kareiva P, G Daily, T Ricketts, H Tallis, S Polasky (eds) Natural Capital: Theory and Practice of Mapping Ecosystem Services. Oxford University Press, New York.

Balmford, A, J Beresford, J Green, R Naidoo, M Walpole, A Manica. 2009. A global perspective on trends in nature-based tourism. PLoS Biology 7: e1000144.

Hill, GW, PR Courtney. 2006. Demand analysis projections for recreational visits to countryside woodlands in Great Britain. Forestry 79: 18-200.

Loomis, JB, RB Richardson. 2006. An external validity test of intended behavior: comparing revealed preference and intended visitation in response to climate change. Journal of Environmental Planning and Management 49: 621-630.

Loureiro, ML, F Macagno, PA Nunes, R Tol. 2012. Assessing the impact of biodiversity on tourism flows: an econometric model for tourist behaviour with implications for conservation policy. Journal of Environmental Economics and Policy 1: 174-194.

Mills, AS, TN Westover. 1987. Structural differentiation: a determinant of park popularity. Annals of Tourism Research 14: 486-498.

Naidoo, R, WL Adamowicz. 2005. Biodiversity and nature-based tourism at forest reserves in Uganda. Environment and Development Economics 10: 159-178.

Neuvonen, M, E Pouta, J Puustinen, T Sievänen. 2010. Visits to national parks: effects of park characteristics and spatial demand. Journal for Nature Conservation 18: 224-229.

Puhakka, L, M Salo, IE Sääksjärvi. 2011. Bird diversity, birdwatching tourism and conservation in Peru: a geographic analysis. PLoS One 6: e26786.

Richardson, R, JB Loomis. 2005. Climate change and recreation benefits in an alpine national park. Journal of Leisure Research 37: 307-320.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Russell, R, AD Guerry, P Balvanera, RK Gould, X Basurto, KM Chan, S Klain, J Levine, J Tam. 2013. Humans and nature: how knowing and experiencing nature affect well-being. Annual Review of Environment and Resources 38: in press.

Trainor, SF, RB Norgaard. 1999. Recreation fees in the context of wilderness values. Journal of Park and Recreation Administration 17: 100-115.

Uyarra, MC, AR Watkinson, IM Côté. 2009. Managing dive tourism for the sustainable use of coral reefs: validating diver perceptions of attractive site features. Environmental Management 43: 1-16.

Williams, ID, NV Polunin. 2000. Differences between protected and unprotected reefs of the western Caribbean in attributes preferred by dive tourists. Environmental Conservation 27: 382-391.

Willox, AC, SL Harper, JD Ford, K Landman, K Houle, V Edge. 2012. "From this place and of this place:" climate change, sense of place, and health in Nunatsiavut, Canada. Social Science and Medicine 75: 538-547.

Wood, SA, AD Guerry, JM Silver, M Lacayo. 2013. Using social media to quantify nature-based tourism and recreation. Scientific Reports 3: 2976.

World Travel and Tourism Council. 2012. Travel and Tourism: Economic Impact.

5.9 Wave Attenuation & Erosion Reduction: Coastal Protection

5.9.1 Summary

Understanding the role that nearshore habitats play in the protection of coastal communities is increasingly important in the face of a changing climate and growing development pressure. The InVEST Erosion Protection model quantifies the protective benefits that natural habitats provide against erosion and inundation (flooding) in nearshore environments. It is composed of two sub-models: a Profile Generator and a Nearshore Wave and Erosion model. In the absence of local data detailing the profile of the nearshore elevations, the Profile Generator model helps you combine information about the local bathymetry and backshore to generate a 1-Dimensional (1D) cross-shore (perpendicular to the shoreline) beach profile. The Nearshore Waves and Erosion model uses the cross-shore profile (either uploaded or created using the Profile Generator) to compute summaries of nearshore wave information, and outputs the total water level at the shore, the amount of shoreline erosion, and the amount of avoided damages due to erosion (in your local currency) from a given habitat management decision that affects the amount of nearshore marine habitats (e.g., coral or oyster reefs, vegetation, sand dunes) at your site. This information can help coastal managers, planners, landowners and other stakeholders understand the coastal protection services provided by nearshore habitats, which can in turn inform coastal development strategies and permitting. The model, of course, has some limitations (see *Limitations and Simplifications*), however, all the science that went into this model is based on well-established models that have been developed and successfully tested by various scientists at many sites, and is expected to be useful for a wide range of management decisions.

5.9.2 Introduction

The Erosion Protection model estimates, by way of a 1D bathymetry transect (or a series of transects) perpendicular to the shoreline, the relative contributions of different natural habitats in reducing nearshore wave energy levels and coastal erosion. These habitats dissipate wave energy and/or act as barriers against high waves and high water levels, and eventually protect coastal properties and communities. The protective services offered by those habitats to coastal populations are highlighted by linking coastal storms to erosion, and then linking erosion to damage of coastal properties. To measure these services, the model has you specify a baseline habitat and a proposed scenario. The proposed scenario can incorporate habitat losses, gains, or some combination of the two over the area of interest. The service provided by these habitats can be measured by the amount of avoided erosion or inundation, by the number of people protected, or by the value of avoided property damages.

The Erosion Protection model is composed of two sub-models: a Profile Generator and a Nearshore Waves and Erosion model. The purpose of the Profile Generator is to assist in the preparation of a 1D bathymetry transect for use in the Nearshore Waves and Erosion model. The inputs of the Profile Generator include the site's location, a file describing the overall shape of the shoreline, and a nearshore topography and bathymetric file that contains land elevation as well as water depths near the site of interest. Furthermore, the model requires information about sediment size, tidal range and backshore characteristics, which is to be input in the Erosion Protection Excel Table (see *Erosion Protection Excel Table*). If this information is not available, we provide guidance on how to approximate these inputs. Outputs of the Profile Generator model include a 1D bathymetry profile at your specified location, information about the site's backshore and the location of natural habitats along the cross-shore transect from offshore to the uplands. In addition, the model provides over-water fetch distances (the distances over which wind blows over water to generate waves) as well as estimates of wave height and wind speeds that can occur at the site of interest during a storm. Overall, this model generates the inputs you need to run the Nearshore Waves and Erosion model.

The Nearshore Waves and Erosion model uses information about the type and location of natural habitat at your site, as well as the way in which your scenario changes those habitat characteristics, to produce estimates of the amount of avoided shoreline erosion or scour and the associated change in property damages. The model inputs are a 1D bathymetry profile (obtained from the Profile Generator model or a site survey) and a value for offshore wave height and period (or a value of wind speed, fetch and the average water depth) that are representative of a certain storm conditions that also have to be described. In addition, the model requires information about the backshore as well as the type and physical characteristics of the natural habitats that are at the site of interest. You will need to specify how your scenario, or management action, will affect the natural habitats footprint or change their density, as compared to the baseline conditions. Model outputs are profiles of wave height under baseline and proposed scenario conditions. The model also estimates the amount of avoided shoreline retreat of a sandy beach or the volume of sediment eroded from consolidated beds (e.g., scour of mud bed) due to your management actions. When you select the valuation option, the model uses the biophysical outputs of the Nearshore Wave and Erosion model coupled with your input about local property value to provide you with information about the change in property damage associated with your scenario.

With this model you can answer a variety of questions. The most basic answer it provides is the incremental change in erosion and property damages from erosion due to a habitat management action. This is simply the net differences in damages between the baseline and proposed scenario. You can also use this model to answer questions about the total level of erosion reduction service provision by habitats in an area. This could be approximated linearly from the incremental value derived from your habitat management action model run. Alternatively, you could propose a scenario where no service provision is provided by habitat and estimate the total value of habitats in that way. Refer to the limitations and simplifications section below for more information on how to interpret these results.

This model should be thought of as estimating one part of the ecosystem services associated to habitats: protection from erosion. When conducting a cost-benefit analysis, it is important to evaluate all value changes arising from a proposed management action. We recommend you run the Coastal Vulnerability model. The Coastal Vulnerability model maps regions that are more or less vulnerable to erosion and inundation during storms and also highlights important characteristics of the region of interest. In addition, it maps regions of the shoreline that are exposed to or sheltered from the open ocean and estimates wind-generated wave characteristics. Coastal Vulnerability model outputs that explore the effects of various management actions (e.g., the presence vs. absence of natural habitats) will help identify regions where natural habitats or a certain management action may have significant impacts on the stability of the coastline. However, the Coastal Vulnerability and Coastal Protection models are independent. You do not need to run the Coastal Vulnerability model in order to run the Coastal Protection model.

5.9.3 The Model

The InVEST Erosion Protection model is a 1D process-based tool that produces an estimate of wave attenuation and erosion reduction owing to the presence of natural habitats, as well as an estimate of the value of those habitats in your local currency unit. A single model run operates along a single transect perpendicular to the shoreline; multiple runs can be distributed along broader swaths of coastline to explore the protective services of natural habitats and the effects of various management actions on the hazards of erosion and flooding within larger regions. At this point, however, the model does not batch process multiple runs so each transect is run on an individual basis.

How it Works

As waves travel from the deep ocean to coastal regions with shallower waters, they start to interact with the seabed. They first increase in height before breaking and dissipate most of their energy in the surf zone and on the beach face. Natural habitats play an important role in protecting shorelines against wave action because they increase the amount of wave dissipation, or, in the case of sand dunes, serve as a physical barrier.

To estimate the profile of wave height that one would expect at a certain region as the wave propagates shoreward three types of information are required:

1. Offshore wave characteristics: wave height and wave period at the deepest point in the bathymetry profile.

2. Nearshore bathymetry and backshore characteristics: elevation **relative to Mean Lower Low Water (MLLW)** of both the submerged (underwater) and emerged (above water) portions of the cross-shore profile.
3. Location and physical characteristics of natural habitats: distance from the shoreline of the natural habitats that will become submerged during a storm, as well as representative density, height and diameter of the habitat elements.

The InVEST Erosion Protection model is composed of two sub-models. The first model, the Profile Generator, helps you obtain cross-shore transect containing nearshore bathymetry and topography information at your site. The Profile Generator can also place the footprints of the natural habitats along that transect. Using this cross-shore profile (or one that you upload), the Nearshore Waves and Erosion model computes profiles of wave height and wave-induced mean water level for your baseline and post-management action scenario. This scenario generally affects the footprint and/or density of the various habitats that you can have at your site. Currently, the model is suitable to value the following sub-tidal (always submerged), inter-tidal (between high and low tides) and supra-tidal (above the high-water mark) habitats: seagrass beds, marshes, mangroves or coastal forests, coral reefs and oyster reefs. The remainder of this section will describe how both the Profile Generator and the Nearshore Waves and Erosion models work.

Profile Generator Model

In order to run the Nearshore Wave and Erosion model, it is necessary to have nearshore bathymetry (water depths) and topography (land elevation) information, as well as the location and characteristics of natural habitats at your site. Also, you must provide an offshore wave height and associated period values. The purpose of the Profile Generator model is to help you glean this information from your site data and help you prepare to run the Nearshore Wave and Erosion model. Additionally, the Profile Generator helps you estimate those data (nearshore elevations and slopes) if you do not have them but know the general characteristics of the site.

First, the Profile Generator helps you obtain bathymetry information by three different options. The model interface asks: “Do you want us to cut a cross-shore transect in GIS?”. If you have a seamless topography/bathymetry (topo/bathy) Digital Elevation Model (DEM) or a bathymetric DEM, you should answer *“Yes”*, and upload a Digital Elevation Model (DEM). If the uploaded DEM is a seamless DEM with both bathymetry and topography represented, the Profile Generator will capture both topographic as well as bathymetric information. Otherwise, if the DEM only captures bathymetric elevations and excludes land elevations above the water level, the cut profile will apply erroneous values of 100.0 (meters) for the missing terrestrial portions. You will have the opportunity to create an ideal backshore profile by filling appropriate information in the Erosion Protection Excel Table (see below).

When you choose the option “Do you want us to cut a cross-shore transect in GIS?”, the Profile Generator works by drawing a transect perpendicular to the shoreline of your site of interest, of a length that you specify in the model interface. This length should be such that a sufficiently deep point is reached and that any adjacent land features are excluded. The model reads the bathymetric and, if the DEM provided is seamless, topographic information along that transect. If the site is surrounded by land (sheltered), or is fronted by an island, the offshore portion of the profile might include the adjacent land feature where waves cannot propagate. To avoid this situation, the model removes any portions of the profile offshore of the deepest point that is shallower than the average depth along the profile.

Another option is to answer *“No, but I will upload a cross-shore profile”*, and then to upload a profile obtained from another source for further processing. At the very least, two coordinate points (X-cross-shore distance from shoreline, and Z-elevation relative to Mean Lower Low Water [MLLW]) in the uploaded file are required. Lastly, if you do not have bathymetric information at the site of interest, you can choose the third option *“No, please create a theoretical profile for me”*, and the model will generate, **for sandy systems only**, a theoretical bathymetric profile, based on the average sand size at the site. The depth profile follows the equation (Dean and Dalrymple, 2002, Chap. 7):

$$Z = AX^{2/3} \tag{5.30}$$

where (X, Z) represent the cross-shore distance and depth, with $X = 0$ at the shoreline. The coefficient A is a profile scale factor and is a function of sediment size (Dean and Dalrymple, p.162 and CEM). This shape of the bed profile is called an equilibrium beach profile, and corresponds to the average profile that one would obtain after averaging years of regular bathymetric surveys at a sandy beach. It can also be viewed as a profile that develops when destructive and constructive forces are in equilibrium. Usually, this profile extends to what's called the "closure depth", which is the depth where waves no longer affect sediment movement on the bottom. However, for simplicity, this profile is extended from the water line down to -20 meters. Please remember that this option is only valid for sandy systems, where sediment size varies between 0.1 to 1.09 mm. Further it is most applicable to oceanic or exposed shorelines (recall, **T0: Coastal Vulnerability** helps designate exposed versus sheltered coastlines).

Once the method that will be used to create an initial bathymetry profile is selected, the Profile Generator will help you modify or add to the information contained in that transect in order to represent the site as accurately as possible. This is especially useful if you want to estimate the amount of erosion at your site and important backshore details are not captured in the elevation (DEM) or habitat (Polygon Shapefiles) inputs. The Profile Generator will also help you modify, smooth or remove portions of the profile to represent the effects of a management action under consideration (e.g., remove offshore portions that are too deep to affect wave heights or remove certain bathymetric features before conducting the analysis).

In order to best use the Profile Generator, it is important that you are familiar with some terminology and typical values of backshore slopes for different environments. Figure 1 shows profiles of a typical beach and a coastal mangrove forest. After waves have propagated from deep water and broken in the nearshore, they reach the foreshore and/or backshore portion of the beach, or, if the water level is high enough, propagate through a marsh or a mangrove forest. Under normal conditions, for sandy beaches there is a relatively flat region between the Mean Lower Low and Mean Higher High (MHHW) water marks called the foreshore. The backshore (the region above MHHW) consists of a berm and, in temperate regions mostly, a sand dune. Berms can range in width from 10's of meters to having a very small or no width. In general, foreshore and backshore information cannot be obtained during standard hydrographic surveys. Also we have found that, although most DEM files have relatively good bathymetric information, intertidal and backshore elevations are often incorrect, unless they were measured during a detailed topographic survey effort. Mangrove forests are usually fronted by a tidal flat with an average slope of 1:1000 to 1:2000, and usually have a relatively monotonic profile whose slope varies between 1:200 to 1:600 (de Vos, 2004; Burger, 2005). In case you would like to measure foreshore and backshore profiles at your site, you can either use standard surveying methods, or follow the simple method in *Appendix A*. However, if you cannot conduct such a survey, you can use the recommendations provided in this guide.

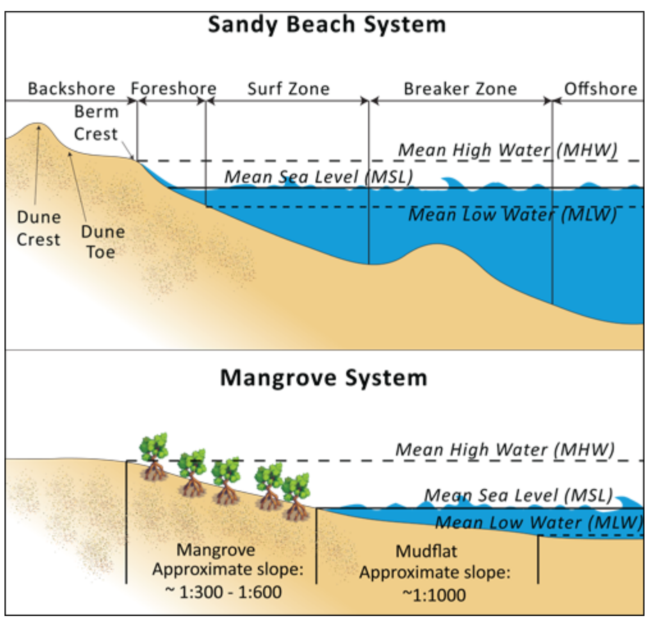


Fig. 5.6: Typical profiles of a sandy beach (top subplot) and a mangrove forest (bottom subplot). Please note the locations of the foreshore in the sandy beach as well as the slope of the mangrove forest.

In the Erosion Protection Excel Table, which summarizes the pertinent characteristics of the profile, you must indicate whether the profile of interest is a sandy beach or a muddy backshore. This option determines what modifications may be made to the cut or user-defined topo/bathy profile. **Option 1. Add backshore to a sandy beach**: assuming that this information is not contained in the cross-shore profile that was cut by the model or in the profile you uploaded, the Erosion Protection Excel Table (see *Erosion Protection Excel Table*) helps you guess what the foreshore slope, berm height and dune height might be for the site of interest, based on simple rules of thumb. Please bear in mind that conditions at the site of interest can differ quite substantially from these rules. Therefore, the suggestions provided should be used as a starting point but a site survey (even as basic as field notes from visual observations) is strongly encouraged if you are interested in obtaining more accurate results.

The average sediment size is required to help approximate foreshore slopes, in case you do not have it. It also helps creating a bathymetric profile for sandy beaches in case you do not have any site measurement. Finally, it is used in the erosion model of the Nearshore Waves and Erosion model. As mentioned earlier, the foreshore is the intertidal region of the beach profile and is assumed to be linear in the model. To provide guidance on what that slope might be, five different values of slope, based on the sediment size, are provided. The first three are derived from observations presented in Wiegel (1964) at beaches that are protected, moderately exposed or fully exposed to the open ocean, in the U.S. The fourth value is derived from observations by McLachlan and Dorvlo (2005) at various beaches around the world. The fifth value is the average of the four previous values. If you do not have the precise sediment size, you can select a value based on a qualitative description of the sand (very fine, fine, medium, course, or very course).

Berm height and foreshore slope often change as a function of seasonal wave climate. After a storm, the profile is flatter and the berm is lower than during fair weather conditions. However, in case you do not have any information about berm height at the site, it is recommended that you place the berm at least at 1 meter above the MHW mark. Finally, a dune height value is needed. Dunes are fairly common in temperate climates and height estimates can be derived from site surveys. In case you do not have this information, we recommend that you enter a value of 2 meters in order to get an estimate of how dunes can protect your site.

Option 2. Add a backshore to a mangrove or marsh. Mangrove and marsh beds are different from sandy beaches because they consist, in general, of consolidated materials, do not have dunes, and their profile is fairly linear. As mentioned earlier and shown in Figure 1, mangrove forests are usually fronted by a tidal flat with an average slope of 1:1000 to 1:2000, and have a relatively monotonic profile whose slope varies between 1:200 to 1:600 (de Vos, 2004; Burger, 2005). If this option is selected, you can enter a maximum of three linear slopes that can be added to the bathymetry profile that you cut/created or uploaded.

The Profile Generator locates the presence of natural habitats along the cross-shore profile. If Option 1 *"Yes"* to the question *"Do you want us to cut a cross-shore transect in GIS?"* is selected, you can also indicate the types of natural habitats that are present in the region of interest, and the model will locate and plot where those habitats fall onto the cross-shore transect. This is done by providing the path to the directory containing separate polygon shapefiles representing the footprints of each habitat. Please note that the results for the habitat placement are accurate only if the natural habitat and bathymetry layers are properly geo-referenced. You should scrutinize results to make sure that the natural habitats are properly placed along the profile (e.g., make sure that seagrass beds are in subtidal areas, or mangroves are in inter- or supra-tidal areas).

Finally, if you do not have any storm wave or wind information at the site to run the Nearshore Waves and Erosion model, the Profile Generator will help you obtain those data by reading and providing you with some pre-processed statistics from the closest WAVEWATCH III (WW3, Tolman (2009)) grid point. Because wave data can be scarce in most regions of the world, 7 years of WW3 model hindcast reanalysis results have been analyzed to estimate, for model grid points that are in waters deeper than 50m, the maximum as well as the average of the top 10% and 25% wave height. The same statistics for wind data, for 16 equiangular direction sectors (0deg, 22.5deg, 45deg, etc.) have also been calculated.

Wind information can be used in the Nearshore Waves and Erosion model by combining it with fetch distance (the distance over which waves are generated by wind) as well as the average depth offshore of the site to compute an offshore wave height and period. The model can compute these fetch distances if you choose *Yes* to the question *Do you want the model to compute fetch distances?*. In that case, from the site location, the model draws 16 equiangular sectors, and in each sector, the model draws nine equiangular radials. Each radial is initially 50km long, but is cutoff when it intersects with a land mass. To capture the effects of those land masses that limit fetch distance, the average fetch distance F_k for each 22.5deg sectors k is weighted by each radial distance and angle (Keddy, 1982):

$$F_k = \frac{\sum_{n=1}^{9} f_n \cos\theta}{\sum_{n=1}^{9} \cos\theta} \tag{5.31}$$

where f_n is the n^{th} radial distance in the k^{th} equiangular sector, and $\theta = 2.5deg$ (22.5deg divided by 9).

From wind speed, and fetch distance, wave height and period of the locally generated wind-waves are computed for each of the 16 equiangular sectors as:

$$\begin{cases} H = \widetilde{H}_\infty \left[\tanh\left(0.343\widetilde{d}^{1.14}\right)\tanh\left(\frac{2.14.10^{-4}\widetilde{F}^{0.79}}{\tanh(0.343\widetilde{d}^{1.14})}\right)\right]^{0.572} \\ T = \widetilde{T}_\infty \left[\tanh\left(0.1\widetilde{d}^{2.01}\right)\tanh\left(\frac{2.77.10^{-7}\widetilde{F}^{1.45}}{\tanh(0.1\widetilde{d}^{2.01})}\right)\right]^{0.187} \end{cases} \tag{5.32}$$

where the non-dimensional wave height and period $\widetilde{H}_\infty$ and $\widetilde{T}_\infty$ are a function of the average wind speed values U that were observed in a particular sector: $\widetilde{H}_\infty = 0.24U^2/g$, and $\widetilde{T}_\infty = 7.69U^2/g$, and where the non-dimensional fetch and depth $\widetilde{F}_\infty$ and $\widetilde{d}_\infty$ are a function of the fetch distance in that sector F_k and the average water depth in the region of interest $d[m]$: $\widetilde{F}_\infty = gF/U^2$, and $\widetilde{T}_\infty = gd/U^2$. $g[m/s^2]$ is the acceleration of gravity. This expression of wave height and period assumes fetch-limited conditions (USACE, 2002; Part II Chap 2). Hence, model results may over-estimate wind-generated wave characteristics at a site if the duration of wind steadily blowing in a fetch direction is less than the time required to realize fetch-limited conditions. Also, wind-waves are not appropriate representations of wave climate on exposed, oceanic coasts. For oceanic coasts, estimates of representative oceanic wave forcing should be used (from WW3 data or another source) rather than wind-wave estimates.

Once a satisfactory bathymetry and topography profile and realistic wave parameters are obtained, you can run the wave Nearshore Waves and Erosion model.

Nearshore Waves and Erosion

The amount of shoreline retreat at sandy beaches is a function of the total water level at the site and storm duration. Conversely, the erosion of muddy shorelines is a function of wave forces on the bed and storm duration. The total water level at the shoreline is composed of the sum of storm surge, wave runup, tide, amount of sea-level rise and any water surface elevation anomaly (e.g., super-elevation during an El Niño). To quantify the protective services provided by natural habitats, the Erosion Protection model computes the amount of attenuation of waves and the reduction in wave-induced mean water level increase (runup) at the shoreline caused by your scenario, as well as the difference in wave-induced velocity at the bed. This information is translated into an avoided erosion amount based on your scenario input, as well as a change in the protective service value of habitats, expressed in your local currency unit.

Wave Evolution Model The first step in this model is to estimate the waves that will "attack" the shoreline. Assuming that waves have a deep water height of H_o and a period T, it is possible to compute the evolution of wave height from offshore to the shoreline along the x-axis of cross-shore transect that you defined with the following wave energy equation:

$$\frac{1}{8}\rho g \frac{\partial C_g H^2}{\partial x} = -D \tag{5.33}$$

where ρ is the density of seawater, taken as $1,024kg/m^3$, $g = 9.81m/s^2$ is the gravitational acceleration, H is the wave height representative of the random wave field, C_g is the speed at which wave energy travels, and D represents the dissipation of wave energy. The role of dissipation is to decrease the amount of wave energy as it propagates through or over different media. It is the sum of the dissipation caused by wave breaking D_{Break}, bottom friction D_{Bot}, and submerged vegetation D_{Veg} :

$$D = D_{Break} + D_{Veg} + D_{Bot} \tag{5.34}$$

Dissipation due to breaking is modeled using the formulation and default parameters presented by Alsina and Baldock (2007), which performed well when compared to various field measurements, even without calibration (Apostos et al., 2008):

$$D_{Break} = A\frac{H^3}{h}\left[\left(\left(\frac{H_b}{H}\right)^3 + \frac{3H_b}{2H}\right)\exp\left(-\left(\frac{H_b}{H}\right)^2\right) + \frac{3\sqrt{\pi}}{4}\left(1 - erf\left(\frac{H_b}{H}\right)\right)\right] \tag{5.35}$$

where erf is the Gauss error function, h is the local water depth, A is the sediment scale factor (see *Profile Generator Model*), and H_b is the maximum wave height prior to breaking:

$$H_b = \frac{0.88}{k}tanh\left(\gamma\frac{kh}{0.88}\right) \tag{5.36}$$

where k is the wavenumber, the ratio of length between two wave crests (called wavelength) L to 2π, and γ is a calibration parameter called the breaking index. The breaking index value, γ, used in the model is the value proposed by Battjes and Stive (1985):

$$\gamma = 0.5 + 0.4\tanh\left(33\frac{H_o}{L_o}\right) \tag{5.37}$$

where H_o and L_o are the deepwater wave height and wavelength, respectively.

The other dissipation terms in Equation (5.34) are expressed as a function of the characteristics of the natural habitats that are present along the profile of interest. In the model, as waves move into portions of the profile with natural habitat, this dissipation term is included. Any non-linear processes that might occur as waves move from one medium or habitat to another are ignored in the model.

Dissipation due to the presence of vegetation is expressed by (Mendez and Losada, 2004):

$$D_{Veg} = \frac{1}{2\sqrt{\pi}}\rho N d C_d\left(\frac{kg}{2\sigma}\right)^3\frac{\sinh^3 k\alpha h + 3\sinh k\alpha h}{3k\cosh^3 kh}H^3 \tag{5.38}$$

where N is the density of vegetation (stems per unit area), d is the frontal width or diameter of vegetation stems, and α represents the fraction of the water depth h occupied by vegetation elements of average stem height h_c: $\alpha = \frac{h_c}{h}$. In the case of emergent vegetation ($h_c > h$), a maximum of $\alpha = 1$ is applied.

Finally, C_d is a taxa-specific (e.g., eelgrass, marsh, mangroves) drag coefficient. Default values of drag coefficient (see e.g., Kobayashi et al., 1983; Bradley and Houser, 2009; Burger, 2005) are applied in the model:

- For seagrass beds and marshes, $C_d = 0.01$
- For trees, including mangroves, $C_d = 1$

For trees, and mangroves in particular, we assumed that roots, trunk and canopy contribute independently to the total dissipation caused by vegetation, and D_{Veg} becomes: $D_{Veg} = D_{Roots} + D_{Trunk} + D_{Canopy}$.

In addition to dissipation caused by vegetative elements, waves can also lose energy because they propagate over a rough bottom such as a coral reef top. Dissipation due to bottom friction is generally initiated when waves are in shallow enough water to "feel" the bottom, and is higher for coarser bed material than smoother ones. In the model, it is triggered when waves travel over sandy bottoms, as well as coral reefs, which are rougher than sand beds. Following Thornton and Guza (1983), the dissipation due to bottom friction is modeled as:

$$D_{Bot} = \rho C_f\frac{1}{16\sqrt{\pi}}\left[\frac{\sigma H}{\sinh kh}\right]^3 \tag{5.39}$$

where C_f is the bed friction coefficient, which is a function of the roughness (or dimensions) of the bed, and σ is the wave frequency, the ratio of wave period T to 2π. In the model, the following default friction coefficients have been assumed:

- For live corals, $C_f = 0.2$,

- For dead (smooth) corals that are still structurally stable : $C_f = 0.1$
- For corals that are structurally compromised and sandy beds: $C_f = 0.001$,

The wave-evolution equation (Equation (5.33)) is valid when the bottom slope is not too steep. When waves encounter a steep barrier such as a coral reef, the model does not compute the amount of breaking dissipation and the profile of wave height during breaking. However, the value of the broken wave height at the edge of the reef top H_r is estimated assuming that wave height is controlled by water depth h_{top} (Gourlay, 1996a, b) : $H_r = 0.46 h_{top}$, where $h_{top} = h_r + \overline{\eta}_r + h_+$ is the total water depth on top of the reef.

The total water depth is the sum of the depth on the reef top referenced to Mean Sea Level h_r, the wave setup on the reef caused by breaking waves $\overline{\eta}_r$, and any additional super-elevation of the water level $\overline{\eta}_+$, which can be caused by tides, pressure anomalies, etc. The wave setup on the reef top is caused by the release of wave energy during breaking and it is computed using the empirical equation proposed by Gourlay (1996a,b; 1997):

$$\overline{\eta}_r = \frac{3}{64\pi} K_p \frac{\sqrt{g} H_i^2 T}{\left(\overline{\eta}_r + h_r\right)^{3/2}} \tag{5.40}$$

where H_i is the incident wave height, or the wave height at the offshore edge of the coral reef. The coefficient K_p is the reef profile shape factor, and is a function of the reef face slope α_f or the reef rim slope α_r, depending on whether waves break on the reef face or rim. Once the broken wave height is established following the equation presented above, the profile of wave height over the reef top is determined following Equation (5.33), with D_{Bot} as defined in Equation (5.39).

Similar to coral reefs, when waves encounter a steep barrier such as an oyster reef, the amount of breaking dissipation is not computed. Instead, the model estimates the wave height H_t immediately shoreward of the reef with the following equations based on the incident wave height H_i immediately offshore of the reef:

$$H_t = K_t H_i \tag{5.41}$$

where K_t is a transmission coefficient. In the case of trapezoidal-shaped reefs, the transmission coefficient is computed with an empirical formula developed for low-crested breakwaters (van der Meer et al., 2005):

$$K_t = \begin{cases} -0.4\frac{R_c}{H_i} + 0.64\left(\frac{B}{H_i}\right)^{-0.31}\left(1 - e^{-0.5\xi}\right) & \text{if } B/H_i < 8 \\ -0.35\frac{R_c}{H_i} + 0.51\left(\frac{B}{H_i}\right)^{-0.65}\left(1 - e^{-0.41\xi}\right) & \text{if } B/H_i > 12 \end{cases} \tag{5.42}$$

where B is the crest width of the reef, and $R_c = h_c - h$ is the crest freeboard, the difference between the reef height h_c and the water depth h. The breaker parameter ξ is computed as $\xi = \tan\alpha / \left(S_i\right)^{0.5}$ where the seaward slope of the reef $\tan\alpha$ is computed as a function of the structure crest and base width, B and W, respectively:

$$\tan\alpha = \frac{2h_c}{W - B} \tag{5.43}$$

Finally, S_i is the incident wave steepness:

$$S_i = \frac{2}{pi} \frac{H_i}{gT_p} \tag{5.44}$$

In the above equation, when $8 < B/H_i < 12$, K_t is estimated by a linear approximation.

If the oyster reef is a dome, the model applies empirical equation proposed by Armono and Hall (2003):

$$K_t = 1.616 - 4.292\frac{H_i}{T^2} - 1.099\frac{h_c}{h} + 0.265\frac{h}{W} \tag{5.45}$$

Once waves have travelled past the coral and oyster reefs, the evolution in the remaining portion of the bathymetry is modeled using the wave evolution equation (Equation (5.33)). It is assumed that the peak period T does not change.

Nearshore Bed Erosion The next step is to model the response of the shoreline to wave attack. The model estimates two types of shoreline response. In sandy beach systems, the amount of shoreline retreat that takes place after a storm is approximated based on the value of storm surge that you input, and the value of wave runup computed by the wave evolution model. When the shoreline is composed of consolidated sediments (mangroves, marshes), the model estimates an hourly amount of bed scour and computes the volumetric sediment loss based on scour rate and storm duration. In both cases, empirical equations are used that ignore the dynamic feedback that takes place between wave and bed as the erosion occurs.

Wave runup (R_2; see USACE (2002, Chap. 4)) is an estimate of the maximum shoreward distance that waves can reach on inundated lands. Once the profile of wave height has been computed, the amount of wave runup at the shoreline is estimated based on the empirical equation proposed by Stockdon et al. (2006):

$$R_2 = 1.1\left(0.35m\sqrt{H_oL_o} + 0.5\sqrt{0.563m^2H_oL_o + 0.004H_oL_o}\right) \tag{5.46}$$

where m is the foreshore slope, or the average cross-shore slope at the shoreline. In the above equation, the first term in the parenthesis represents the wave setup, and it can be influenced by the presence of the vegetation. The second term represents the wave swash, and it is composed of two terms. The first term, which is a factor of the foreshore slope m is called incident wave swash, and it can also be influenced by the presence of the vegetation. The second term is the called the infragravity swash. It is assumed that this term is not affected by the presence of vegetation elements because vegetation does not affect long-period waves as much as it does short period waves (Bradley and Houser, 2009). In the absence of biogenic features, the Erosion Protection model only requires information on the characteristics of offshore waves and foreshore slope to compute wave runup with Equation (5.46). If intertidal or subtidal biogenic features are present, wave runup is estimated via a series of steps described below.

First, the wave height profile is estimated, in the absence and in the presence of vegetation, following the procedure outlined above. From these wave height profiles, the wave setup $\overline{\eta}$ at the shoreline is estimated by solving the following force balance equation:

$$\frac{\partial S_{xx}}{\partial x} + \rho g\left(h+\overline{\eta}\right)\frac{\partial \overline{\eta}}{\partial x} - f_x = 0 \tag{5.47}$$

where S_{xx} is the force per unit length generated by the waves on the water column, and f_x is the force per unit area due to the presence of vegetation elements:

$$f_x = -\alpha F_x \tag{5.48}$$

where the force F_x is computed following Dean and Bender (2006):

$$F_x = \rho g \frac{1}{12\pi} N d C_d \frac{k}{\tanh kh} H^3 \tag{5.49}$$

Neglecting non-linear processes associated with wave propagation, this equation is only valid for emergent vegetation. Consequently, the coefficient α is added to approximate the effects of vegetation on the wave setup when it is submerged. This approximation over-estimates the reduction in wave setup caused by submerged vegetation compared to what would be obtained if a non-linear wave theory to estimate F_x were adopted. However, this approximation is much faster and simpler to adopt.

Once a value of wave setup in the absence of vegetation has been obtained, a proportionality coefficient β between the empirical estimate of wave setup and the value of the modeled wave setup at the shoreline $\overline{\eta}_{Shore}$ is computed:

$$\beta = \frac{\overline{\eta}_{shore}}{0.35m\sqrt{H_oL_o}} \tag{5.50}$$

Based on the modeled value of the wave setup at the shoreline in the presence of vegetation, $\overline{\eta}^v_{Shore}$, the hypothetical offshore wave height H_p that would have achieved the same modeled setup is computed, assuming that the value of the coefficient β is the same:

$$H_p = \frac{1}{L_o}\left(\frac{\overline{\eta}^v_{Shore}}{0.35m}\right)^2 \tag{5.51}$$

In cases when the effects of vegetation are so pronounced that $\overline{\eta}^v_{Shore}$ is negative, it is assumed that $H_p = 0$. We adopted this empirical approach as a way to estimate the way in which vegetation affects runup, in the absence of observations or models.

Finally, to estimate the amount of runup at the shoreline in the presence of natural habitats, H_o is replaced in Equation (5.46) by the value of the hypothetical offshore wave height H_p in the wave setup and wave-induced swash terms:

$$R_2 = 1.1\left(0.35m\sqrt{H_pL_o} + 0.5\sqrt{0.563m^2H_pL_o + 0.004H_oL_o}\right) \tag{5.52}$$

where the last term is left untouched because, as mentioned earlier, it has been assumed that long waves are not affected by the presence of natural habitats. Similarly, the value of the offshore wavelength L_o is not changed because it has been assumed that peak wave period is not affected by the presence of natural habitats.

From the value of runup at the shoreline, the amount of beach retreat (sandy berm) or volumetric sediment loss (mud) can be computed. Sandy beaches are eroded during storms and generally build back during periods of fair weather. The amount of shoreline erosion is a function of the elevations of sand berm and dunes in the backshore, the wave height and period during the storm, the length of the storm and the total water level elevation during the storm.

As mentioned earlier, the total water level during the storm is a function of the storm surge elevation, wave runup elevation, the tide stage during the storm and any super-elevation of the water surface caused by large-scale oceanic processes (e.g. El Nino). In the model, a storm surge elevation value is required as input and as well as offshore (starting) wave height and period. From these forcing inputs, the model computes the amount of runup for the different management actions that you wish to evaluate from Equation (5.46). Consequently, it is important that you adjust the bathymetry profile to any other water surface elevation difference that you wish to evaluate in the model. For example, if you are interested in investigating wave inundation and erosion at high tide, the elevation of high tide should be added to the value of the surge for a given storm.

The distance of sandy beach retreat during a storm E_s is estimated following the model proposed by Kriebel and Dean (1993):

$$E_s = -\frac{1}{2}(1-\cos\alpha)E_\infty \tag{5.53}$$

where the beach potential erosion response if the storm lasted an infinite amount of time E_∞ is scaled by the duration of the storm under consideration by a time-correction factor α. The potential erosion response E_∞ is computed as a function of the wave breaking characteristics and the backshore dimensions:

$$E_\infty = \frac{S(x_b - h_b/m) - W(B + h_b - 0.5S)}{B + D + h_b - 0.5S} \tag{5.54}$$

where S is the total water level during the storm, referenced to MSL (please note that the model adjusts the bathymetry to MSL based on the tide information that you provided in the Erosion Protection Excel Table, so **the initial bathymetry profile should be referenced to MLLW**). h_b and xb represent the water depth and distance from the shoreline where the offshore wave breaks with a height H_b. Breaking wave characteristics are computed by applying the wave evolution equation, Equation (5.33), to an equilibrium profile built from the sediment scale factor corresponding to the sediment size at the site (see Equation EqProf in *Profile Generator Model*). E_∞ is also a function of the foreshore slope m, as well as the height and width of the sand berm B and W, and dune height D in the backshore, as well as, the specified berm height, B, and breaking depth, h_b. Equation (5.54) is only valid up to a certain maximum surge elevation. E_∞ becomes erroneously negative or undefined if:

$$B + h_b \leq \frac{S}{2} \tag{5.55}$$

If this condition arises, the model incrementally adds 0.5 meters to the berm height B until (5.55) is untrue. The beach retreat E_∞ associated with this adjusted berm height is computed rather than using the height that you provided. The output report produced by the model will notify you that the berm height has been adjusted and by how much if this is the case.

The scale coefficient α ($\pi \leq \alpha \leq 2\pi$) is computed by solving the following equation:

$$\exp(-\alpha/\beta) = \cos\alpha\breve{}(1/\beta)\sin\alpha \tag{5.56}$$

where β is a function of the finite storm duration T_d and breaking wave characteristics:

$$\beta = 320\frac{2\pi}{T_d}\frac{H_b^{3/2}}{\sqrt{g}A^3}\left(1 + \frac{h_b}{B+D} + \frac{mx_b}{h_b}\right)^{-1} \tag{5.57}$$

Practically, the model estimates the amount of beach retreat that would occur under various management scenarios by first solving Equation (5.53) in the absence of vegetation. Breaking location is computed as explained above, using the sediment scale factor A derived from the sediment size that you input. In the presence of vegetation, it is often difficult to estimate the exact location of breaking, and there is not any guidance or observation of avoided beach retreat in the presence of natural habitats. Consequently, the amount of beach retreat in the presence of natural habitats is estimated by scaling the amount of retreat obtained in the absence of natural habitats by the ratio of reduction in runup values as well as the ratio of the cube of wave height over the submerged vegetated bed. This is because empirical models of beach retreat are directly proportional to water level (e.g., see Equation (5.54)). Also, process-based models of beach erosion (e.g., Kriebel and Dean, 1985) scale erosion by wave dissipation, which is proportional to the cube of wave height. The model's final output value of erosion in the presence of natural habitat is the average of both values.

Note: You may notice that for certain values of m, Equation (5.54) can yield negative results. Instead of generating a message error, the profile foreshore slope is decreased so that E_∞ is positive. This correction is made because of the uncertainty associated with the erosion model and your inputs. In future versions of this model, a more sophisticated erosion model will be used to avoid this situation. To estimate a correct foreshore slope that won't yield negative values in Equation (5.54), the model approximates the breaking wave height by using Equation (**??**) (see *Profile Generator Model*). Then the model computes the breaking position and depth x_b and h_b by assuming that $H_b = 0.78h_b$ and:

$$h_b = Ax_b^{2/3} \tag{5.58}$$

If the model does adjust the profile slope, be cautious of comparing retreat values to values obtained at neighboring locations or at the same site for other forcing conditions. An increase in slope causes an increase in retreat not associated with increased forcing or the lack of protective habitats.

In addition to sandy beaches, the model can also estimate the volumetric erosion a consolidated bed might experience. Muddy substrates, such as those found in marshes or mangrove forests, do not erode in the same manner as sandy beaches. They are composed of cohesive sediments that are bound by electro-magnetic forces, and their resistance to wave- and storm-induced bed velocity is a function of their composition and level of consolidation. In the erosion model, the hourly rate of scour of a consolidated bed $E_m[cm.h^{-1}]$ is estimated by following the method proposed by Whitehouse et al. (2000, Ch. 4):

$$E_m = \begin{cases} 36(\tau_o - \tau_e)m_e/C_M & \text{if } \tau_o - \tau_e > 0 \\ 0 & \text{if } \tau_o - \tau_e \leq 0 \end{cases}$$

where m_e is an erosion constant and C_M is the dry density of the bed. Both constants can be obtained from site-specific measurements. However, the Erosion Protection Excel Table offers sample default values of $m_e = 0.001m.s^{-1}$ and $C_M = 70kg.m^{-3}$. The variable τ_e is the erosion shear stress constant (the maximum shear stress the consolidated bed can withstand before sediment begins to scour) and is computed as:

$$\tau_e = E_1 C_M^{E_2} \tag{5.59}$$

where E_1 and E_2 are site specific coefficients. The erosion threshold value within the model has be prescribed using average values of those coefficients (Whitehouse et al., 2000): $E_1 = 5.42 \cdot 10^{-6}$ and $E_2 = 2.28$. Finally, the wave-induced shear stress τ_o is computed as:

$$\tau_o = \frac{1}{2}\rho f_w U_{bed}^2 \tag{5.60}$$

where U_{bed} is the wave-induced bottom velocity at water depth h:

$$U_{bed} = 0.5H\sqrt{g/h} \tag{5.61}$$

and f_w is the wave-induced friction coefficient, computed assuming the flow is turbulent:

$$f_w = 0.0521\left(\frac{\sigma U_{bed}^2}{\nu}\right)^{-0.187} \tag{5.62}$$

where $\nu \approx 1.17 \cdot 10^{-6} m^2.s^{-1}$ is the kinematic viscosity of seawater, and $\sigma = 2\pi/T$ is the wave frequency.

The model estimates the rate of bed erosion for regions that are above MLLW, assuming that there is no mixture of sand and mud in the inter- and supra-tidal areas. Since the wave height H and, therefore velocity at the bed U_{bed} decays from the shoreline moving inland, the model is able to compute the spatial variation of the scour rate with respect to distance from the shoreline. By integrating under the spatially varying scour rate curve and multiplying by the duration of the storm, the model also yields an approximate of the volumetric sediment loss at along the modeled profile. The model also returns the distance inland where erosion is expected based on the inland limit of where the bed shear stress exceeds the threshold value. Further, since the reduction in habitat footprint and/or density will increase wave heights and, therefore, scour rates, the model computes the spatially varying scour rates and volumetric sediment loss for the present and modified habitat footprints. In other words, the model estimates the increase in erosion due to the removal or modication of natural habitats.

Valuation The Erosion Protection model quantifies the protection provided by habitats in terms of the avoided damages to property due to erosion from waves. The market value of properties in the area, based on a sample of recent sales, can be used as an estimate of property value. Tax assessment data and replacement cost methods are the two other common ways to quantify the value of properties. Any of these three options are valid inputs to this model. As this model is intended to work in data-poor regions, you only need to provide information about average property value in a given area. However, the extent to which this will reflect the true value of these properties, and thus the value of habitat in providing protection from storms, directly depends on the quality of the property value input. For more information on the merits of each data source and a general discussion of valuation using this approach, please refer to Cannon (1995).

Coastal storms damage properties, and the difference in damages due to habitat can be considered an indicator of the value of those habitats in providing protection to properties from storms. The main inputs to the valuation module are the areas of erosion from the baseline and management scenarios, avoided land loss amount, and local data on property values that you provide. The difference in distance eroded between the baseline and management scenarios is referred to as “avoided erosion” (R_A) or avoided land loss:

$$R_A = R_2 - R_1 \tag{5.63}$$

where R_1 and R_2 are estimates of the total amount of land area eroded under habitat scenario one (baseline) and habitat scenario two (management scenario) over a longshore distance L over which outputs of the waves and erosion are valid. We leave it up to you to define this longshore distance L. However, in general, it can be defined as an area where bathymetry, topography and natural habitat characteristics do not vary much. The values of land eroded R_x under each habitat scenario, $x = 1, 2$, is obtained from the erosion outputs of the erosion model E_x as:

$$R_x = E_x L \tag{5.64}$$

As changes in land use need to be considered against other possible investments and time preferences, it is appropriate to consider the expected present value, EPV, of services provided by habitat. The EPV calculation employs a discount rate, i, over a time horizon that you define, τ, expressed in years. EPV reflects the value of the stream of avoided storm damages over time due to a change in habitat and discounts the value of those avoided damages in distant periods when the discount rate is greater than zero. We have provided a default discount rate, but you should assess whether that is appropriate for your case. For more information, see this website. The EPV is also a function of the expected return period associated with your storm. Storms are classified by their strength and probability of occurrence. Thus, it is common to hear them referred to as a “hundred-year storm” or a “thousand-year storm,” where the expected frequency of a storm of that strength is once per hundred or thousand years. The annual probability of occurrence for a storm that occurs on average every T years is $p = 1/T$, where p is constant across time - that is, not contingent on previous occurrences.

The model estimates the value of habitats for coastal protection from erosion as the difference in damages under two habitat scenarios, given as $D_A = D_2 - D_1$ for a given storm class with an expected return time of T. Because storms occur at irregular intervals over time, the model allows you to assess these benefits across a defined time horizon. EPV for a given storm class is calculated as:

$$EPV = \sum_{t=1}^{\tau} \frac{pD_A}{(1+i)^t} \tag{5.65}$$

5.9.4 Limitations and Simplifications

Although the coastal protection model will help you inform management decisions by demonstrating the protective capabilities of natural habitats, it has limitations (theoretical and otherwise). A primary limitation is that the Erosion Protection model assumes that all erosion leads to a loss of land. In some places this assumption will reflect reality; in other locations, erosion from large storms can be reversed through net sediment accretion during periods of calm weather. These sorts of more complex physical dynamics are beyond the scope of the model and analysis we present here. Further, the model estimates coastal protection services provided by habitats in terms of the reduction in damages due to erosion from storm waves, not surge. Some coastal habitats have the ability to attenuate surge in addition to waves (e.g., marshes, coastal forests), while other nearshore subtidal habitats do not (e.g., eelgrass). If you are modeling storm waves from hurricanes which also produce significant surge, the current model likely underestimates the protection value of the former habitats.

In addition to the limitations discussed the above, the model has technical limitations. The first is the lack of high quality GIS data that are readily available. In the event that you do not have a nearshore profile for the region of interest, simple rules of thumb based on observations are provided to help you generate one. Though grounded in the literature, these rules of thumb will not generate profiles that perfectly match all sites of interest. Again, a site visit to obtain missing data will improve the generated profile, and thus the model results.

The theoretical limitations of the Nearshore Waves and Erosion model are more substantial. As mentioned earlier, wave evolution is modeled with a 1D model. This assumes that the bathymetry is longshore-uniform (i.e. the profile in front of the site is similar along the entirety of the stretch of shoreline). Because this is unlikely true, the model ignores any complex wave transformations that occur offshore of the site of interest. Also, although the wave model used compares well against observation with default calibration parameters (see *Nearshore Waves and Erosion*) you are not currently offered the option to calibrate it. Thus, values of wave height and wave-induced water level along the modeled transect might differ from observations.

Another limitation of the wave model is that it has been assumed that the vegetation characteristics that you provide in the Erosion Protection Excel Table remain valid during the storm forcing that is being modeled. The model also ignores any non-linear processes that occur when waves travel over submerged vegetation. For example, the model does not take into account wave reflection that occurs at the edge of the vegetation field, motion of vegetative elements caused by wave forces, or reductions in habitat density that might occur during a storm. Furthermore, default values of friction and drag coefficient are used to compute the forces exerted by the habitats on the water column. This implies that those forces are independent of the flow turbulence regime. Finally, simple empirical models are used to compute the wave profile over coral and oyster reefs. Although these models have been validated with observations, they ignore many processes that might change the wave profile that the model computes. You should also be aware that, while under some small levels of storm surge oyster reefs provide some wave protection, the primary role of oyster reefs is to prevent wave erosion of saltmarshes during typical or day to day wave conditions and water levels.

To model beach erosion, the model proposed by Kriebel and Dean (1993) is used. Although this empirical model has been widely used (USACE, 2002), it ignores key erosion processes that occur during a storm. For example, the dynamic response and feedback between waves and the bed profile during the storm is not taken into account. The model also does not evaluate when dune breaching and the amount of overwash that might occur during the simulated storm.

To model scour of consolidated beds, the model proposed in Whitehouse et al. (2000) is used, and, in the Erosion Protection Excel Table, default sediment characteristics are provided but are not appropriate for all sites. Further, the assumption that the whole bed has the same characteristics, both horizontally and vertically, is made. Finally, any dynamic response between increase levels of suspended sediments and wave-induced bottom velocity, as well as any sediment redeposition and settlement, are ignored. Site-specific input parameters might help improve the accuracy of model results relative to using the provided default parameters, but will not compensate for the physical simplifications made.

The avoided damages model can be categorized as a partial equilibrium analysis. It investigates the erosion protection value of habitat with a process-based model by changing only the level of habitat, holding all else constant. However, it is important to realize that a dramatic change in habitat management may decrease the reliability of results. For example, you could make the case that property values would change in response to a large shift in habitat, invalidating our assumption that these values stay constant in the baseline and proposed scenario.

The expected reduction in erosion due to habitats calculated for the model is for a storm of the size modeled in the wave evolution and nearshore erosion modules. However, during the time period defined by you, the coastal habitats will likely provide protection against a wide range of different sized waves and storms. The most accurate way to value habitats for their total protection services would be conduct multiple runs of the model for different sized storms that occur with different return periods (e.g., different hurricane categories) and add together the expected avoided damages due to habitat protection outside of the model, such that:

$$EPV_{AllStorms} = \sum_{t=1}^{\tau_1} \frac{p_{Storm1} D_{A1}}{(1+i)^t} + \sum_{t=1}^{\tau_2} \frac{p_{Storm2} D_{A2}}{(1+i)^t} + \sum_{t=1}^{\tau_2} \frac{p_{Storm2} D_{A2}}{(1+i)^t} + \ldots \tag{5.66}$$

You can perform this calculation by adding together the results from multiple runs of the InVEST model for different sized storms with different return periods.

Finally, this model does not explicitly account for property owners' behavioral response to erosion over time. Depending on this response, the assessed value may overstate or understate the potential damages. As an example of this behavioral response, communities may erect manmade erosion protection, property owners may raise their houses, many may sell their property and hence property values will decrease, etc.

In summary, the interactions between waves and the shoreline represent extremely complex processes. Calculating avoided damages from erosion using data that aggregates property value (e.g., land value and structure value) into one value simplifies the diverse and complex ways individual properties are damaged by erosion (e.g., loss of land versus damages to structures). Thus, our estimates of the value of coastal protection services do not distinguish between these possibilities. The simple model presented here is designed to capture the essence of these and to guide your understanding of the roles that nearshore habitats might play in mitigating the coastal hazards of erosion and inundation.

5.9.5 Data Needs

As mentioned earlier, the Erosion Protection model is composed of two primary sub-models: the Profile Generator and the Nearshore Waves and Erosion models. It is recommended that you first utilize the Profile Generator tool to obtain a cross-shore profile that contains bathymetry and backshore information. This tool will also help you obtain several pieces of useful information including: the bathymetry and nearshore topography along the profile of interest; the type of natural habitats present at the site, as well as your location along the profile; values for offshore wave height and wind speed and fetch direction for the site. Once this profile information has been obtained and forcing parameters have been selected, you can run the Nearshore Waves and Erosion model. Also, to investigate the impacts of management actions on waves and erosion, you can select the type of management action or change the footprint and density of each habitat. Running the Nearshore Waves and Erosion model requires, at a minimum, a bathymetry profile as well as wave and storm information. Furthermore, information on the type of backshore present at the site, as well as on the characteristics of the natural habitats that are present at the site will be needed.

Profile Generator

1. **Workspace (required).** You need to specify a workspace folder path where the model outputs can be stored. It is recommended that you create a new folder that will contain all the model outputs (Profile Generator as well as Nearshore Waves and Erosion outputs). For example, by creating a folder called "WCVI" inside the "Coastal-Protection" folder, the model will create "_Profile_Generator_Outputs" and/or a "_NearshoreWaveErosion" folders containing outputs from your various runs, as well as an intermediate folder named "scratch".

```
Name: Path to a workspace folder.  Avoid spaces.
Sample path: \InVEST\CoastalProtection\WCVI
```

2. **Label for Profile Generator Run (10 characters max) (required).** Provide a short name that reflects the location or reason of your run. This name will be used to create a subfolder inside the "_Profile_Generator_Outputs" folder that will contain outputs for your model runs. For example, if you chose the label "Dune_2m" because you wanted to see what a cross-shore profile with a 2m dune looked like, a folder called "Dune_2m" inside the "_Profile_Generator_Outputs" folder will be created. That folder will contain two subfolders called "html_txt" and "maps". The "html_txt" folder contains an html file that summarizes information about the site of interest with figures of the created profile and showing the location of natural habitats along the profile. The "maps" folder contains shapefiles that can be viewed in GIS. These shapefiles include polylines that show fetch vectors and fetch distances, points along the transect where topo/bathy was extracted as well as points showing the locations of natural habitats.

```
Name: A concise label describing the model run
File type: text string (direct input to the ArcGIS interface)
Sample: Dune_2m
```

3. **Land Point (required).**. Provide a point shapefile of the location where you want to run the Profile Generator. The datum of this input must be WGS 1984 in order to avoid transformations when the model projects the Wave Watch III Model Data input. It is highly recommend that you use snapping to ensure that the point is on the edge of the land polygon (shoreline). From this location the Profile Generator will extract a profile orthogonal to the land (if you are cutting a transect in GIS), gather wind and wave data from the closest deep-water WW3 grid point, and/or compute fetch distances, averaged over 16 directions. **If you are cutting a cross-shore transect in GIS, make sure to inspect the coastline around this input and adjust the Land Point Buffer Distance (input 8) accordingly.**

```
Name: File can be named anything, but no spaces in the name
File type: point shapefile (.shp)
    Sample path (default): \InVEST\CoastalProtection\Input\LandPoint_BarkSound.shp
```

4. **Land Polygon (required).** This input provides the model with a geographic shape of the coastal area of interest, and instructs it as to the boundaries of the land and seascape.

```
Name: File can be named anything, but no spaces in the name
File type: polygon shapefile (.shp)
Sample path (default): \InVEST\Base_Data\Marine\Land\LandPolygon_WCVI.shp
```

5. **Do you want us to cut a cross-shore transect in GIS? (required).** This drop down box allows you to select whether you 1) wish to have the GIS create a cross-shore transect, 2) will upload a cross-shore profile of your own or 3) prefer to have the model create a theoretical profile. The answer provided to this question will determine whether subsequent inputs are required or optional.

```
File type: drop down options
Sample: (1) Yes
```

6. **Bathymetric Grid (DEM) (optional).** If you have answered "(1) Yes" to the question: "Do you want us to cut a cross-shore transect in GIS?", the model requires a DEM in order to cut a cross-shore profile. This bathymetric grid layer should have a vertical elevation referenced to Mean Lower Low water.

```
Name: File can be named anything, but no spaces in the name
File type: raster dataset
Sample path: \InVEST\Base_Data\Marine\DEMs\claybark_dem
```

7. **Habitat Data Directory (optional).** If you have answered "(1) Yes" to the question: "Do you want us to cut a cross-shore transect in GIS?", the model will optionally allow for the location of natural habitats that intersect on the cross-shore transect. To do so, you must store all Natural Habitat input layers that you want to consider in a unique directory. Each natural habitat layer should consist of the location of those habitats, and all data in this folder must be polygon shapefiles and projected in meters. Further, each of these layers should end with an underscore followed by a unique number, for example "_1" or "_2". The model allows for a maximum of six layers in this directory. Do not store any additional files that are not part of the analysis in this folder directory. If you need to add or remove natural habitat layers at one site for various analyses, you will have to create one "Natural Habitat" folder per analysis (omitting the habitat you wish to remove). If you wish to exclude natural habitat from your analysis, simply leave this input blank.

```
Name: Folder can be named anything, but no spaces in the name
File type: None, but must contain polygon shapefiles (.shp)
Sample path: \InVEST\CoastalProtection\Input\NaturalHabitat
```

8. **Land Point Buffer Distance.** If you have answered "(1) Yes" to the question: "Do you want us to cut a cross-shore transect in GIS?", the model requires this distance value in order to create a perpendicular transect based upon the slope of the coastline near the Land Point (input 3). The Land Point shapefile must be within this buffer distance from the shoreline as defined by the Land Polygon (input 4). Also, the terrestrial area located behind or in front of that point must be wider than the buffer distance. In general, a distance of 250m is sufficient. However, if the site is along a narrow island or a spit that distance should be smaller than the width of the island or the spit. **It is recommended that if your Land Point is placed near a sinuous coastline (e.g. surrounded by narrow inlets), you should determine the maximum distance from the Land Point in both directions along the coast without crossing an abrupt change in angle of the coastline. This distance measure should be entered as the Land Point Buffer Distance and will allow the model to determine the true angle for a transect perpendicular to this Land Point site.**

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the ArcGIS interface)
Sample (default): 250
```

9. **Length of your profile.** If you have answered "(1) Yes" to the question: "Do you want us to cut a cross-shore transect in GIS?", the model requires the length of the profile you wish to create from the Land Point (input 3) to a suitable offshore limit (in km). If the provided DEM is seamless, the Profile Generator extracts topography for the same length inland of the point. This length should be the distance from the Land Point to the deepest adjacent point (in a sheltered region or in an estuary) such that an adjacent land masses are not intersected, or to a sufficiently deep point along an open or exposed coastline.:

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the ArcGIS interface)
Sample (default): 25
```

10. **Cross-Shore Profile (optional).** If you have answered "(2) No, but I will upload a cross-shore profile" to the question: "Do you want us to cut a cross-shore transect in GIS?", the model will not cut a cross-shore profile for you from a GIS layer, but will create a smooth backshore profile, or manipulate a cross-shore profile of your choice. This file must contain a minimum of 2 (X,Z) coordinates. It must be tab delimited with two columns. The first column must be the cross-shore distance X-axis, where X=0 is at the shoreline (positive X pointing seaward, negative X pointing landward). The spatial resolution of the X-axis (spacing between two X-coordinates) must be equal to 1 (dx=1). The second column must indicate the cross-shore elevations along the X-axis. Depths values must be negative (referenced to Mean Lower Low Water) and terrestrial elevations must be positive.:

```
Name: File can be named anything, but no spaces in the name
File type: Tab delimited text file with two columns (X,Z) (.txt)
Sample path: \InVEST\CoastalProtection\Input\Depths.txt
```

11. **Smoothing Percentage (required).** Enter a percentage value for how much you wish to smooth the profile created or fed through the model. A value of "0" means no smoothing.

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the ArcGIS interface)
Sample (default): 5
```

12. **Erosion Protection Excel Table (required).** This file contains information about your site that will allow the model to build a full cross-shore profile, including tidal elevations, and profile slope modifications. Also, the locations of natural habitats will be populated here by the Profile Generator Model if you include the Habitat Data Directory as input. This table has 4 sections: General Site Information, Foreshore/Backshore Profile Modifications, Habitats, and Habitat Management Action. Three of the sections, General Site Information, Foreshore/Backshore Profile Modifications, and Habitats are applicable to the Profile Generator tool. In the Foreshore/Backshore Profile Modifications section, you have the option of modifying the topo/bathy profile by inserting linear slopes along the profile. You are required to populate the Habitats section if you include a Habitat Directory in the Profile Generator model. For more information on how to complete this Erosion Protection Excel Table, please see *Erosion Protection Excel Table*.

```
Name: File can be named anything, but no spaces in the name
File type: *.xls or .xlsx (if you have MS Excel 2007 or newer)
Sample path: \InVEST\CoastalProtection\Input\ErosionProtection_WCVI_BarkSound.xls
```

13. **Wave Watch III Model Data (optional).** If you would like the model to gather wind and wave statistics that might represent oceanic conditions at your site, upload the WW3 file that has been provide in the InVEST download package. The model will use this dataset to read the maximum, top 10% and top 25% wind speed as well as wave height and associated wave period values from the model grid closest to your site.

```
Name: File can be named anything, but no spaces in the name
File type: polygon shapefile (.shp)
Sample path: \InVEST\CoastalProtection\Input\WaveWatchIII.shp
```

14. **Wave Watch III Search Distance (kilometers).** The model requires this search distance in order to find the closest WW3 point. The default distance is 50 km, but may need to be increased depending on the distance of your Land Point to the nearest WW3 point. To determine the appropriate distance for your site, use ArcGIS to measure the distance (over water) of the Land Point to the nearest WW3 Model Data point.

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the ArcGIS interface)
Sample (default): 50
```

15. **Do you wish to calculate fetch for Land Point? (optional).** This drop down box allows you to specify whether you want the model to compute fetch distances. If "(1) Yes" is selected, fetch radials will be extended from the Land Point (input 3) and cut based on the Land Polygon (input 4). The results will be averaged over 16 directions.

```
File type: drop down options
Sample: (1) Yes
```

Nearshore Waves and Erosion

The Nearshore Waves and Erosion model estimates the profile of wave height over your bathymetry from an offshore value to the shoreline. It is used to estimate the amount of erosion of a beach or a muddy substrate. This section explains how to obtain and/or interpret all the data the model requires to run properly.

1. **Workspace (required).** You need to specify a workspace folder path where model outputs will be stored. It is recommend that you input the same workspace folder that you input in the Profile Generator, which will contain all model outputs (Profile Generator as well as Nearshore Waves and Erosion outputs, see *Profile Generator*). In this workspace, we will create a folder name "_WaveModel_Outputs" that will contain all Nearshore Waves and Erosion outputs.

```
Name: Path to a workspace folder.  Avoid spaces.
Sample path: \InVEST\CoastalProtection\WCVI
```

2. **Label for Waves and Erosion Run (10 characters max) (required).** Provide a short name that reflects the reason for your run. This label will be used as a suffix to all outputs created inside the "_WaveModel_Outputs" folder. For example, if you chose the label "Dune_2m" to evaluate the protective services provided by a 2m sand dune, the model will create an html output file named "OutputWaveModel_Dune2m" as well as a text file indicating wave height as a function of cross-shore distance named "WaveHeight_Dune2m"

```
Name: A concise label describing the model run
File type: text string (direct input to the ArcGIS interface)
Sample: Dune_2m
```

3. **Erosion Protection Excel Table (required).** You are required to fill out and upload the Erosion Protection Excel Table. This table contains information about tide levels, the type of substrate at your site, the type and physical characteristics of natural habitats, and how the management action affects the natural habitats. For more information on how to complete this Erosion Protection Excel Table, please see *Erosion Protection Excel Table*.

```
Table Names: File can be named anything, but no spaces in the name
File type: *.xls or .xlsx (if you have MS Excel 2007 or newer)
Sample: InVEST\CoastalProtection\Input\ErosionProtection_WCVI_BarkSound.xls
```

4. **Cross-Shore Profile (required).** A cross-shore profile is required (which can be obtained from the Profile Generator's outputs) in order to model wave height evolution in your area. The output text file can be found in the "html_txt" folder of a successful PG run and will be called "CreatedProfile_[suffix].txt". This file must contain a minimum of 2 (X, Z) coordinates, and must be tab delimited with two columns. The first column must be the cross-shore distance X-axis, with X=0 at the shoreline (positive X pointing seaward, negative X pointing landward). The spatial resolution of the X-axis (spacing between two X-coordinates) must be equal to 1 (dx=1). The second column must indicate the cross-shore elevations along the X-axis. Depth values must be negative (referenced to Mean Lower Low Water) and terrestrial elevations must be positive.

```
Name: File can be named anything, but no spaces in the name
File type: Tab delimited text file with two columns (X,Z) (.txt)
Sample path: InVEST\CoastalProtection\WCVI\_ProfileGenerator_Outputs\Dune_2m\html_txt\CreatedPro
```

5. **Do you have wave height and wave period values? (required)** The model requires the wave height and period at the offshore edge of your profile as starting conditions. This drop down box allows you to select whether you 1) will provide wave height and wave period values or 2) will instead provide wind speed, fetch distance, and water depth. If you choose answer 1: "Yes, I have these values", enter them below the prompts starting with "IF 1:". If you choose answer 2: "No, please compute these values from wind speed and fetch distance", enter a wind speed, fetch distance as well as average water depth at your site below the prompts starting with "IF 2:". If you have run the Profile Generator and input WW3 data and had the model compute fetch distances for you, you can use that model run's html outputs for default values of wave height and period, wind speed and fetch distances. Figures 12 and 13 can also be used as guidance for typical wave height and wind speed observed during certain classes of storms.

```
File type: drop down options
Sample: (1) Yes
```

6. **Wave Height (meters) (optional).**: Wave height is the distance between the wave crest and wave trough, as shown in the figure under Fetch Distance (below). For typical values of wave period during storms, see the

following figure.

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the ArcGIS interface)
```

Weakly circulating tropical system with winds under 45 mph.		
Tropical storm Circulating tropical system with winds over 45 mph and less than 75 mph.	Very steep seas. Highest waves in squall lines.	H 5 - 8 m T 5 - 9 sec
Hurricane Intense circulating storm of tropical origin with wind speeds over 75 mph. Shape is usually roughly circular.	Can produce large wave heights. Directions near storm center are very short-crested and confused. Highest waves are typically found in the right rear quadrant of a storm. Wave conditions are primarily affected by storm intensity, size, and forward speed, and in weaker storms by interactions with other synoptic scale and large-scale features.	Saffir Simpson Hurricane Scale SS H(m) T(sec) 1 4-8 7-11 2 6-10 9-12 3 8-12 11-13 4 10-14 12-15 5 12-17 13-17 (see Table IV-1-4)
Extratropical cyclones Low pressure system formed outside of tropics. Shapes are variable for weak and moderate strength storms, with intense storms tending to be elliptical or circular.	Extreme waves in most open-ocean areas north of 35° are produced by these systems. Large waves tend to lie in region of storm with winds parallel to direction of storm movement. Predominant source of swell for most U.S. east coast and west coast areas.	Weak: H 3-5m T 5-10 sec Moderate: H 5-8m T 9-13 sec Intense: H 8-12m T 12-17sec Extreme: H 13-18m T 15-20sec

Coastal Engineering Manual, USACE 2002

Fig. 5.7: Typical values of wave height and associated wave period for various types and classes of storms. Use this information to make the best possible guess of wave characteristics offshore of your site.

7. **Wave Period (seconds) (optional).**: Wave period is the amount of time, in seconds, necessary for two consecutive wave crest to pass a fixed point (see the figure under Fetch Distance below). Wave period should be less than 20s. For typical values of wave period during storms, see the preceding figure.

```
Name: A numeric text string smaller than 20 seconds (positive integer)
File type: text string (direct input to the ArcGIS interface)
```

8. **Wind Speed (meters per second) (optional).**: Strong winds blowing steadily over the water can generate high waves if the fetch distance is long enough. Please enter a wind speed value that is representative of the conditions that you want to represent at your site. Please remember that wind patterns at your site might have a seasonal signature and vary depending on the direction they blow towards. If you have uploaded WW3 data in the Profile Generator, we provide you in the html output a wind rose representing typical storm wind speeds at your site, coming from 16 equiangular directions. Also, the following figure can also be used as a guidance for typical wind speed observed during certain classes of storms.:

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the ArcGIS interface)
```

9. **Fetch Distance (meters) (optional).**: Fetch is defined here as the distance travelled by winds over water with no obstructions, for a certain compass direction. Winds blowing over a longer fetch generate higher waves than winds blowing over a smaller fetch distance. You can get fetch directions for the 16 equiangular directions that form a compass by choosing the fetch option in the Profile Generator tool (see the following figure).

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the ArcGIS interface)
```

10. **Water Depth (meters) (optional).**: For a given fetch distance, wind blowing over a shallow area generate

Table IV-1-4				
Saffir-Simpson Damage-Potential Scale				
Scale Number (category)	Central pressure (millibars)	Wind speed (m/sec)	Surge (m)	Damage
1	≥980	33-42	~1.5	Minimal
2	965-979	43-49	~2-2.5	Moderate
3	945-964	50-58	~2.6-3.9	Extensive
4	920-944	59-69	~4-5.5	Extreme
5	<920	>69	>5.5	Catastrophic
(From Hsu (1998); originally from Simpson and Riehl (1981))				
Adapted from Coastal Engineering Manual, USACE 2002				

Fig. 5.8: Typical values of central pressure, wind speed and surge level for various classes of hurricanes. Use this information to make the best possible guess of wind speed offshore of your site if you want the model to estimate values of wind-generated wave height and period during your storm. Also, use this information to make the best possible guess of surge elevation during your storm.

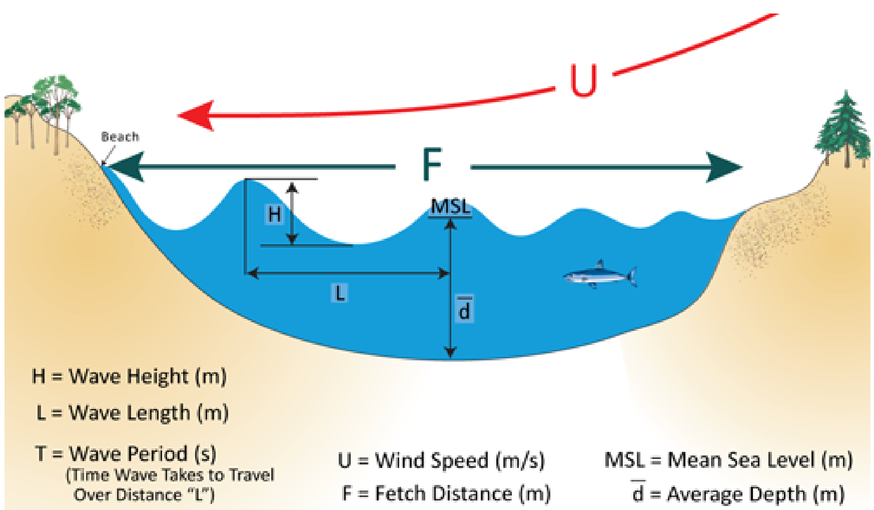


Fig. 5.9: Definition of various coastal engineering terms used in the model.

smaller waves than wind blowing over the deep ocean. Here, enter the average depth value along the fetch angle that you have chosen (see the preceding figure). This value will be used to generate realistic values of wave height and associated period at your site.

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the ArcGIS interface)
```

11. **Storm Duration (hours) (required).**: In order to estimate the amount of beach erosion or bed scour in inter- and/or supra-tidal areas, enter the maximum water level reached during your input storm, as well as its duration. Please indicate the duration of the storm you wish to model.

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the ArcGIS interface)
Sample (default): 5
```

12. **Surge Elevation (meters) (required).**: In order to estimate the amount of beach erosion or bed scour in inter- and/or supra-tidal areas, enter the maximum water level reached during your input storm. Please make sure that the storm surge level you input is consistent with the wind speed or wave height that you entered. For guidance, please consult the Wind Speed figure for storm surge levels typically observed during hurricanes. This surge elevation is applied to the MSL. If you want to investigate, for example, a storm hitting your area at high tide you must add the high tide elevation to this surge value and enter the sum for this input.

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the ArcGIS interface)
Sample (default): 1
```

13. **Model Spatial Resolution (dx) (required)**: A coarse spatial resolution can sometimes lead to model instability and inaccuracy in model outptuts. Please choose a proper resolution at which you want to run the model. This value can be greater or smaller than one. However, keep in mind that a smaller resolution results in longer computing time.

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the ArcGIS interface)
Sample (default): 1
```

14. **Compute Econonomic Valuation (optional)**: By checking this box, users will instruct the model that they would like to approximate a monetary value for habitat and the loss in this value owing to habitat modification (reduction).

15. **Longshore Extent (meters) (required for valuation)**: To obtain an approximate area of land loss associated with retreat/erosion, the retreat/erosion distance must be multiplied by a longshore length. Essentially, this is the length along the shore where one would expect the same amount of retreat. In other words, this is the along shore length where the natural habitat types, coverage, and management actions, as well as, topo/bathy and forcing conditions are approximately uniform.

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the ArcGIS interface)
Sample (default): 250
```

16. **Property Value (local currency) (required for valuation)**: This is the average monetary value of the land, per square meter, that you wish to use in the valuation computation. This model is envisioned as a way to calculate property damages due to erosion, however you could use any source of per-area value. For example, if you had data on the replacement cost of beaches per square meter this model could provide estimates of the avoided replacement costs due to habitat. Value enters into the model as a function of area eroded, so it can be used to measure any valid loss in value due to erosion.:

```
Name: A numeric text string (required for valuation)
File type: text string (direct input to the ArcGIS interface)
```

17. **Return Period of Storm (years) (required for valuation)**: This is the number of years between occurences of the storm (surge and waves) applied in the model run that is experienced at your site. More extreme storms are more infrequent than less extreme storms. Typical return period used in risk assessment are 10, 50, 100, and 500 years, with 10 years being the most common and mild conditions and 500 years being very extreme and infrequent/less likely storm conditions.

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the ArcGIS interface)
Sample (default): 10
```

18. **Discount Rate (required for valuation)**: A discount rate to adjust the monetary benefits of the natural habitats in future years to the present time is required. We provide a default value of 5%, however you are strongly encouraged to evaluate the appropriate rate for your decision-making context.:

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the ArcGIS interface)
Sample (default): 0.05
```

19. **Time Horizon of Valuation (years) (required for valuation)**: This is the number of years over which you intend to value the coastal protection services provided by your habitat.

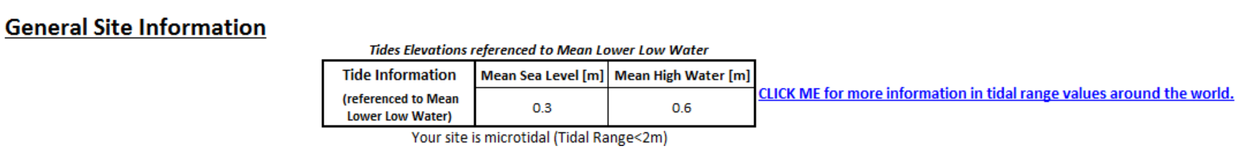
```
Name: A numeric text string (positive integer)
File type: text string (direct input to the ArcGIS interface)
```

Erosion Protection Excel Table

The Erosion Protection Excel Table contains four sections: General Site Information; Profile Modification; Habitats; and Habitat Management Action.

General Site Information

1. Tidal Elevations: You are to enter the elevation of Mean Sea Level (MSL) and Mean High Water (MHW) relative to Mean Lower Low Water (MLLW). Since most bathymetric/nearshore surveys are conducted at the lowest tides, it has been assumed that the vertical datum of the source of bathymetry data (DEM, text file corresponding to an actual cross-section survey, etc.) is MLLW. If it is known that the vertical datum of the bathymetry data is something other than MLLW, enter the elevation of MSL and MHW relative to the known datum. For example, if the vertical datum is actually MSL and the elevation of MHW above MSL is 0.5 m, a value of 0 and 0.5 should be entered in MSL and MHW columns, respectively. In the example shown in the screenshot below, the topo/bathy elevations are presumed to be relative to MLLW, and MSL and MHW are 0.3 m and 0.6 m above MLLW, respectively.


General Site Information

Tides Elevations referenced to Mean Lower Low Water

Tide Information	Mean Sea Level [m]	Mean High Water [m]
(referenced to Mean Lower Low Water)	0.3	0.6

Your site is microtidal (Tidal Range<2m)

CLICK ME for more information in tidal range values around the world.

Fig. 5.10: Screenshot of the Tide Information fields within the "General Site Information" section of the Erosion Protection Excel Table.

This information is used by the Wave and Erosion Model to shift the profile depths to be relative to MSL. Also, a link is provided in the Excel table to a figure showing tidal ranges (the difference between MHHW and MLLW elevations) from around the world. If you are uncertain of the tidal elevation values you have entered, you can check this figure to ensure if the tidal range agrees with the values that you have entered. Otherwise, you can approximate MSL as half the value of the tidal range and MHW as the value of the tidal range shown in this figure. As with all inputs, if accurate local measurements of tides are available, these data should be used.

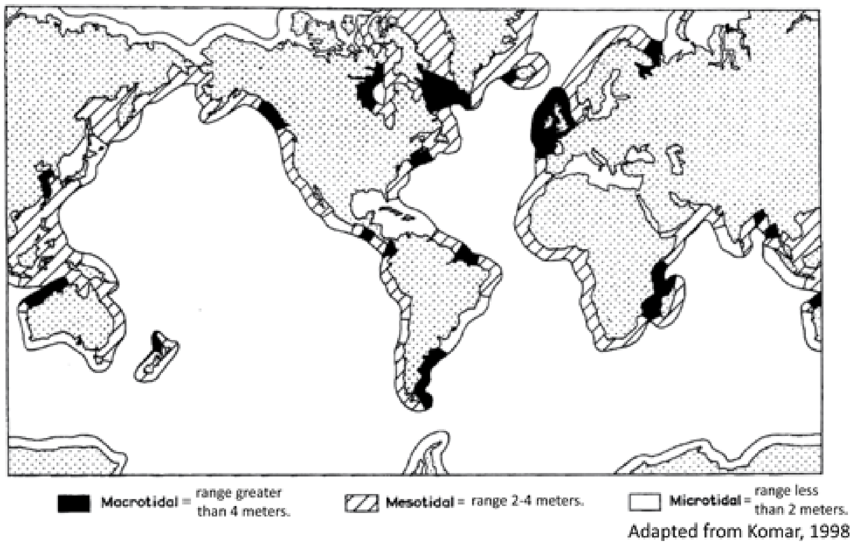


Fig. 5.11: Worldwide variation of tidal range. This information can be used to make the best possible guess of tide elevation at the site of interest.

2. Type of backshore, Sediment and Beach Characteristics: Here, you define what type of sediments make up your backshore. Please refer back to cp-ProfOptions for a more complete description of the two options. Option number 1 corresponds to a sandy backshore and option 2 corresponds to a muddy backshore; this tells the Wave and Erosion Model which erosion computation to run. You must also enter the median diameter or size of the sediment at your site. If you have a qualitative description of the sediment at your site (coarse sand, very fine sand, silty, etc.), a representative sediment size can be obtained by using the Unified Soil Classification (from Dean and Dalrymple, 2002, Ch. 2) shown below; a link to this figure is contained in the Excel table.

If the sediment size does not correspond to the backshore option (if Option 1, sandy beach is selected and the sediment size corresponds to clay/mud, for example), an error message lets you know that you must change the sediment size to agree with the backshore option.

If the option is a sandy beach and a valid sediment size is entered, you are required to enter the following characteristics of your sandy beach: dune height, berm width, berm elevation, and foreshore run. The dune height is the elevation of the dune crest relative to the berm. If you are unsure whether or not dunes exist at your site, a map showing the worldwide distribution of dunes is provided and is shown below.

The berm width is the width of the sandy beach from the shoreline to the toe of the dune or other backshore feature (coastal development, estuary, etc.). The berm elevation is the elevation of the sandy beach relative to MSL. It is recommended that the berm elevation be *at least* as high as the elevation of MHW. Lastly, the foreshore run is the inverse of the foreshore slope. The Excel table populates suggested foreshore runs for you to choose from based on sediment size. The figure below shows a pictorial definition of these characteristics of a sandy backshore.

Since berm height and width, as well as dune elevation, is easily obtained from visual estimates, we encourage you to visit your site to obtain the most accurate values for these parameters. Also, the foreshore slope can be obtained from a simple survey method, see *Appendix A*.

If the option is a muddy system and a valid sediment size is entered, the model requires a dry density value and an erosion constant for the sediment at the site. These parameters cannot be approximated by visual observations or simple methods but require laboratory testing of site samples. Therefore, default values are provided in the Excel table. If you have these values specifically for your site or region, you can overwrite these defaults.

In addition to informing the Wave and Erosion Model about which erosion models to run, and the important physical characteristics for those models, this information also informs the Profile Generator. For example, if a sandy beach is selected, the Profile Generator will incorporate the beach geometry (foreshore slope, berm height and width, and dune

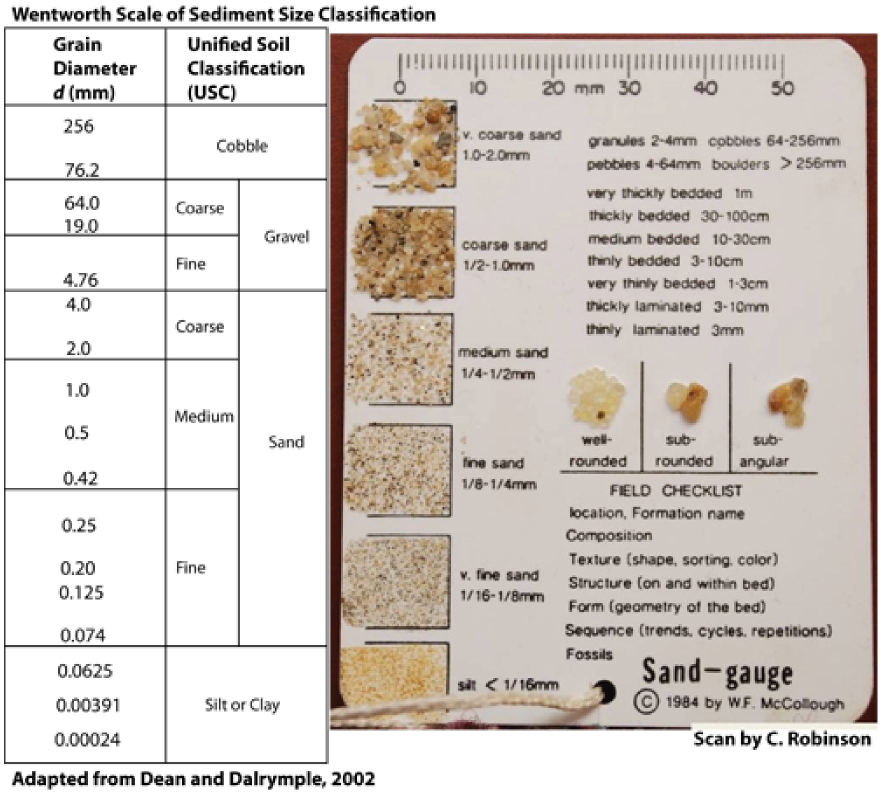


Fig. 5.12: Sediment size classification. Use the table and Geotechnical Gauge to make the best possible guess of sediment size at the site.

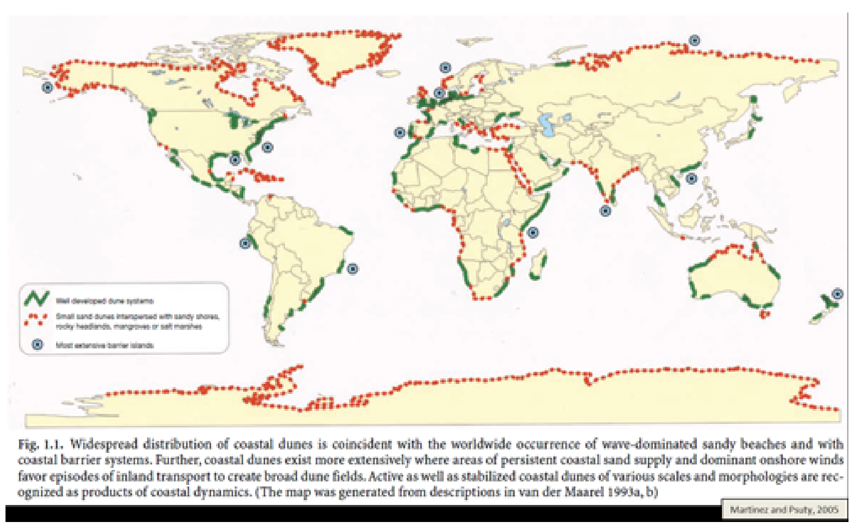


Fig. 5.13: Map showing the approximate distribution of sand dunes in the world. This information can be used to make a guess about whether or not there's a sand dune at the site.

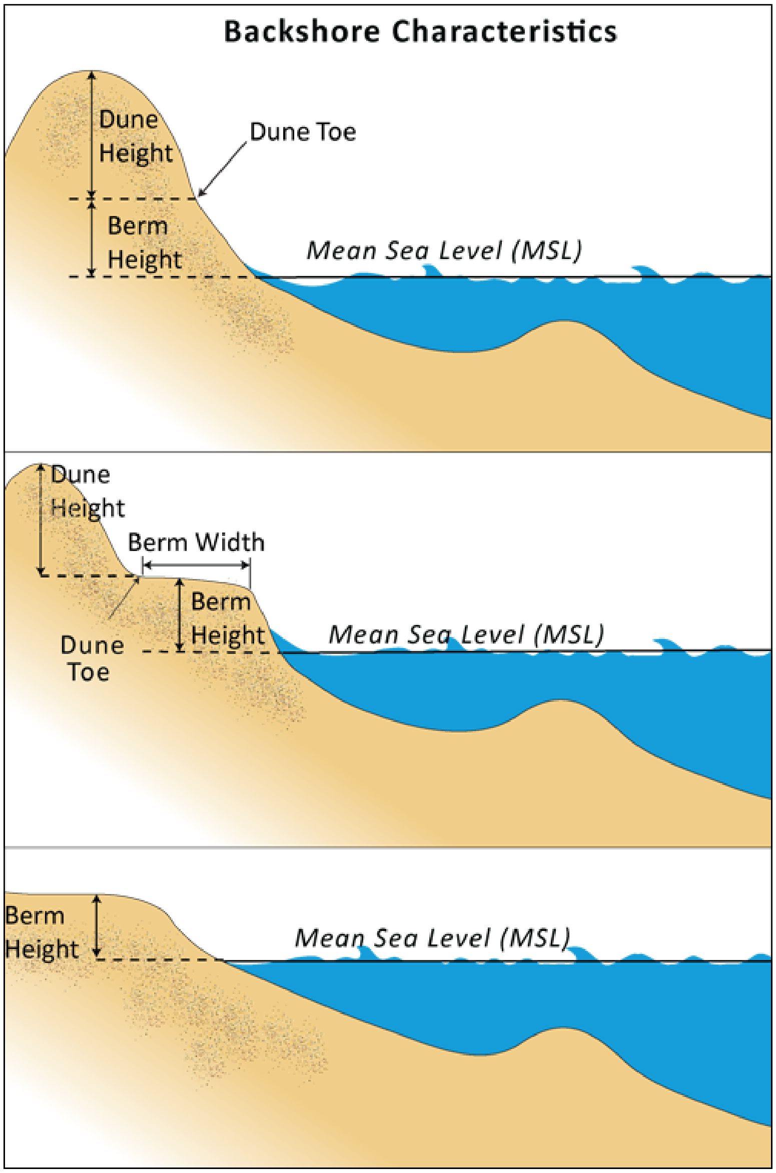


Fig. 5.14: Definition of Berm Height, Berm Width and Dune Height at a typical sandy beach.

height) into the generated profile. Also, if you opt for the Profile Generator to create an Equilibrium Beach Profile (for sandy beaches only), the Profile Generator uses the sediment size provided here to compute the sediment scale factor (see Equation EqProf).

The figure below is a screen capture of where you enter these backshore and sediment characteristics. In the example below, the site is a sandy beach with medium sized sand. Since the option and sediment corresponds to a sandy beach, the dry density and erosion constant fields are greyed out. If this example corresponded to a muddy system, the sandy beach fields would be greyed out and the dry density and erosion constant fields would appear.

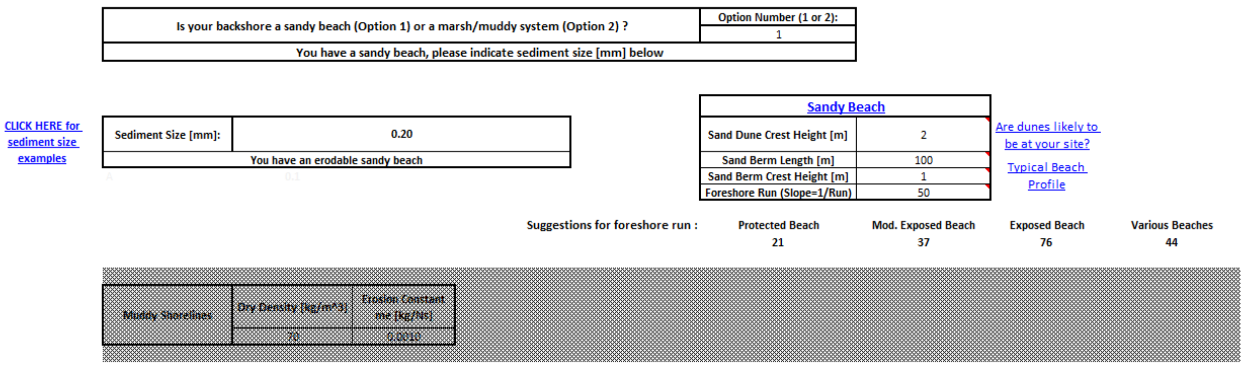
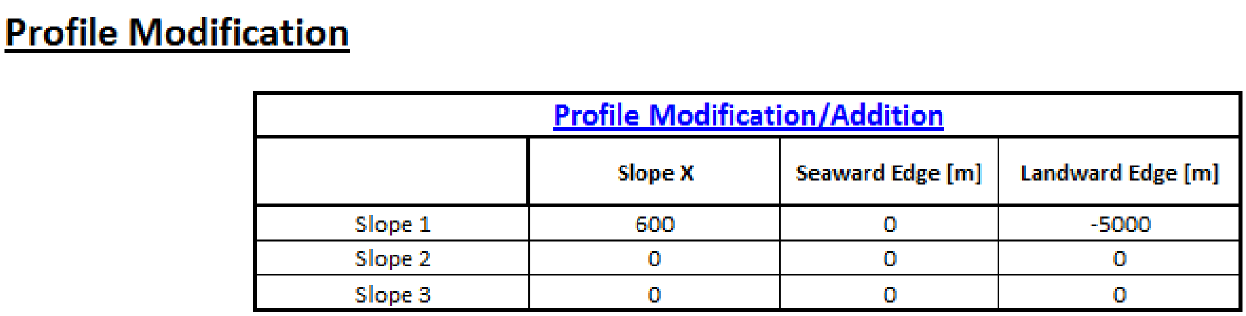


Fig. 5.15: Screenshot of the backshore and sediment characteristic fields within the "General Site Information" section of the Erosion Protection Excel Table.

Profile Modification

In this section, you can superimpose three linear (monotonic) segments onto their topo/bathy profile. To add a monotonic profile, the run value "R" (slope=1/R) as well as the cross-shore locations between which this monotonic slope will apply are required. For a flat profile, you can either enter 0 or a very large number. The convention used is that the beginning point of the transect is seaward of the end point. Also, the origin of the X-axis is at the shoreline, with positive X pointing offshore, and negative X pointing landward of the shoreline. In the screenshot shown below, the user wishes to place a slope of 1/600 from the shoreline to 5 km (5000 m) onshore.



Profile Modification

Profile Modification/Addition

	Slope X	Seaward Edge [m]	Landward Edge [m]
Slope 1	600	0	-5000
Slope 2	0	0	0
Slope 3	0	0	0

Fig. 5.16: Screenshot of the "Profile Modification" section of the Erosion Protection Excel Table.

This example likely corresponds to a case where the elevation was not seamless, or there was no topography measurement, and the user is applying a typical slope associated with mangroves as the backshore profile.

Habitats

In this table, you indicate the types of natural habitats that are present in the Natural Habitats folder that was specified in the Profile Generator prompt. If you intend to have the Profile Generator place habitats on the cross-shore profile rather than record the locations manually, this table must be filled out. To let the Profile Generator know which layer in the folder corresponds to which habitat type, you will need to enter in the Habitat ID cell the number that corresponds to the suffix in the shapefile name corresponding to that habitat (e.g., "1", or "5", etc.). If a particular habitat is not present, those cells should be blank. In the example below, mangroves, seagrass beds, and coral reefs are present in the study region and the suffix corresponding to these habitats are 1, 2, and 3, respectively

Habitats

Enter the suffix number associated with the layers in your Habitats folder. For habitats not in your directory leave the "ID" blank.

Habitat ID	1		2		3	
Habitat Name	Mangroves	Marsh	Seagrass Beds	Sand Dunes	Coral Reef	Other

Fig. 5.17: Screenshot of the "Habitat" section of the Erosion Protection Excel Table.

Habitat Management Action

1. **Sand Dune**: If your management action includes reducing the height of your sand dune (or if you would like to investigate the increase in erosion if your sand dune was lowered or removed), you should enter the percent height reduction in this field. A value of 0 corresponds to no change while 100 corresponds to full removal. In the example shown below, the management action is to reduce the height of the dune by 75%.

Sand Dune	Crest Height Reduction (%)	75

Fig. 5.18: This is where you can define the percent reduction in your dune height associated with a management action.

2. **Vegetation**: You can specify the physical characteristics of three types of nearshore vegetation: mangroves, seagrass and marshes. You can treat coastal forests as mangroves. For each vegetation type, you need to indicate a representative height, stem diameter and stem density. See the following figure for a definition of those terms, and see the next figure for sample values of these characteristics for seagrass, marshes, and mangroves.

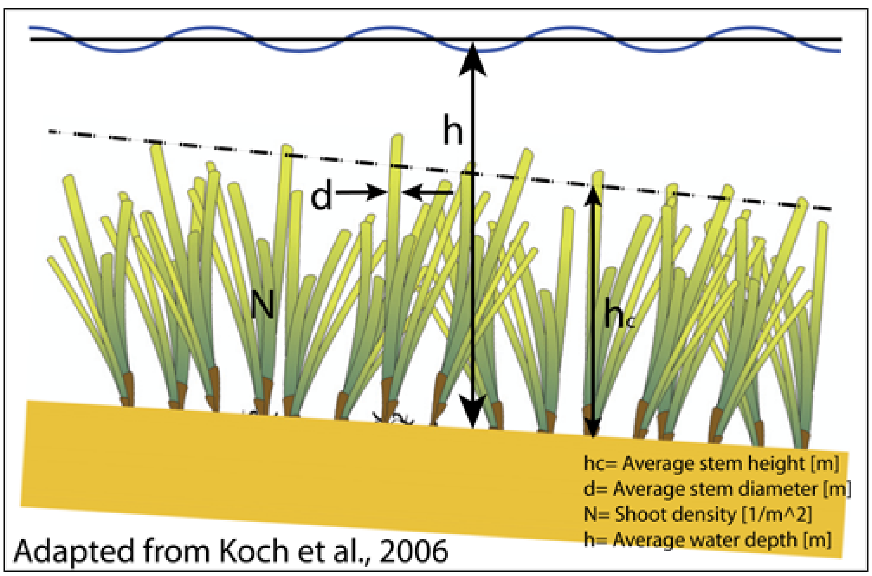


Fig. 5.19: Definition of vegetation characteristic terms used in the model.

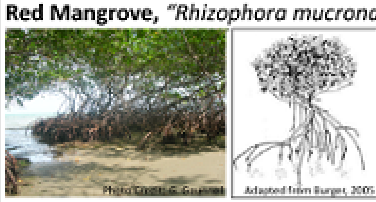
You also need to indicate the distance of their landward and seaward edges from the shoreline (X=0). In our convention, positive X point offshore, and negative X point landward. All vegetation in inter- and supra-tidal regions should have negative X positions and if positive x-locations are assigned for mangroves or marshes, the model assumes that you intended those values to be negative. If you properly included natural habitat in a Profile Generator run, the

Marsh and Seagrass Systems Vegetation Characteristics

	Height [m]	Diameter [m]	Density [#/m^2]
Marsh	0.15-0.35	0.05	2400
Seagrass	0.2-1	0.004-0.075	240-2000

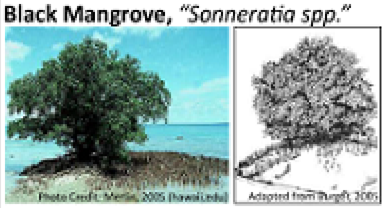
Mangrove Systems Vegetation Characteristics

Red Mangrove, *"Rhizophora mucronata"*



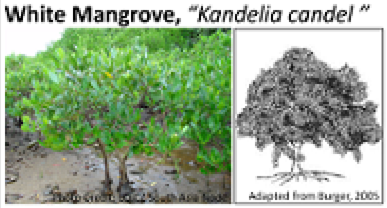
	Height [m]	Diameter [m]	Density [#/m^2]
Roots	0-1	0.05-0.1	1-130
Trunk	5-8	0.15-0.4	0.5-1.7
Canopy	0.2-3	0.02-1	1-100
(Narayan, 2009)			

Black Mangrove, *"Sonneratia spp."*



	Height [m]	Diameter [m]	Density [#/m^2]
Roots	0.1-0.8	0-0.07	4-131
Trunk	0.8-15	0.12-0.5	0.08-1.7
Canopy	0.2-3	0.02-1	1-100
(Mazda et al., 1997 and Narayan, 2009)			

White Mangrove, *"Kandelia candel"*



	Height [m]	Diameter [m]	Density [#/m^2]
Roots	0.13	0.15	1
Trunk	0.17	0.07	1
Canopy	0.5	0.003	100
(Mazda et al., 1997)			

Fig. 5.20: Typical example of vegetation characteristics values for the various habitats used in the model.

Pre-Management Action positions will be populated for you but you should double check these values; the Profile Generator may place marsh or mangrove habitats slightly offshore because of differences in projections, precisions, and accuracy of the input layers. Finally, you will have to indicate how they are affected by your management action:

- You can change the footprint or location of the vegetation. If the vegetation is completely removed, you should have 0's for the X locations post-management action. If the footprint is unaffected, the pre- and post-management action footprints should match.
- You can also change the density of each vegetation type independently. The model will reduce the density of the habitat for the post-management action by the percentage provided.

The following is a screenshot showing the section of the Excel table where the physical characteristics, pre- and post-management locations, and percent density reduction for vegetative habitats are populated. In the example shown, marshes are present from the shoreline (X=0) to 600 meters inland. The marsh footprint is unaffected by the management action but the density is reduced by 20%. There is also a seagrass bed present from 50 to 500m offshore. The post-management location is reduced to between 50m and 400m offshore but the density is unchanged.

			Physical Characteristics			Before Management Action		After Management Action		
			Height [m]	Diameter [m]	Density [1/m^2]	Closest Distance to Shore [m]	Furthest Distance to Shore X [m]	Closest Distance to Shore [m]	Furthest Distance to Shore X [m]	Change in Density [%]
Vegetation	Mangroves	Root System	0.5	0.075	90	0	0	0	0	0
		Trunk	6	0.3	1.2					0
		Canopy	1	0.2	75					0
	Marshes		0.57	0.00057917	3584	0	-600	0	-600	20
	Seagrass Beds		0.5	0.02	540	50	500	50	400	0

Fig. 5.21: A screenshot of the habitat management action section of the Excel table for the vegetation type habitats.

3. **Coral Reef**: If you have a coral reef at your site, we will evaluate the wave height at its shoreward edge based on its dimensions. First, you need to specify its location along the profile that you uploaded as well as the type of the reef that is present:

   - If the reef type is a barrier, enter "-1" for both the offshore and shoreward edge locations and "Barrier" for the reef type.
   - If the reef is located at the shoreward edge of your profile, such as in the case of a fringing reef without a lagoon, the reef location should have the closest distance to shore as 0. The reef type should be defined as "Fringe".
   - If the reef is located somewhere along your profile, with a lagoon on its shoreward edge, please enter its location as accurately as possible. The reef type should be defined as "Fringe Lagoon".

   Second, you need to specify the physical characteristics of the reef, as defined in the following figure: reef face slope, reef rim slope, depth at reef edge, depth on reef top and width of reef top. Most of these data are obtained through site-specific surveys. However, in case you do not have those data, you can still use our model by entering "0" for the reef face slope, the reef rim slope and the depth at reef edge. You can measure reef width from aerial pictures of your site or from global databases of coral reef (see the Coastal Vulnerability model). Finally, you can enter a best guess for reef top depth knowing that reef top depth values vary between 1 and 2 meters, on average. In this case, we will estimate the wave height on the reef top by assuming that waves break on the reef face, and take an average value for the coefficient K_p in Equation (5.40).

   Finally, you need to specify how coral reefs are affected by your management action:

   - If coral reefs are dead but their skeleton is still in place, enter "Dead". In that case, we will reduce the bottom friction coefficient experienced by waves by half (see *Nearshore Waves and Erosion*).
   - If coral reefs are dead and their skeleton failed, enter "Gone". In this case, we will assume that the reef is now a sandy bottom and adjust the bottom friction coefficient accordingly.
   - If the reef is not affected by your management action, enter "None".

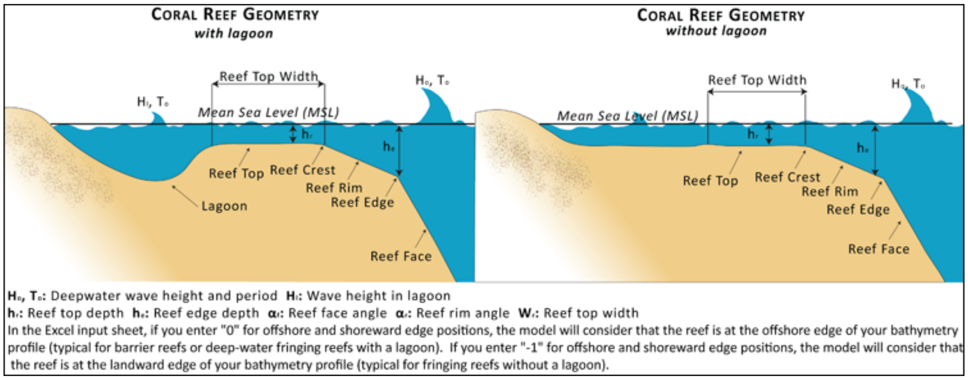


Fig. 5.22: Profiles of coral reefs in the presence or absence of a lagoon, along with definition of the terms used in the Excel input sheet.

In the screenshot shown below, there is a Fringe Lagoon reef type located from 200m to 500m offshore that will be included in the model. The slopes are unknown but the edge depth, top depth, and top width are 10m, 2m, and 230m, respectively. The management action assigned is "Gone".

Coral Reef	Type (Barrier, Fringe, Fringe Lagoon)	Closest Distance to Shore [m]	Furthest Distance to Shore X [m]	Reef Face Slope αf	Reef Rim Slope αr	Depth at reef edge he [m]	Depth on Reef Top hr [m]	Width of Reef Top Wr [m]	Management Action (Gone, Dead, None)
	Fringe Lagoon	200.0	500.0	0.00	0.00	10.0	2.0	230	Gone

Fig. 5.23: An example of inputs for management actions on a Fringe Lagoon reef.

4. **Oyster Reef**: If you have oyster reefs at your site, you need to enter its distance from the shoreline, as well as its dimensions (see the following figure). If you have a Reef Ball $^{(TM)}$, enter "0" for the crest width. :

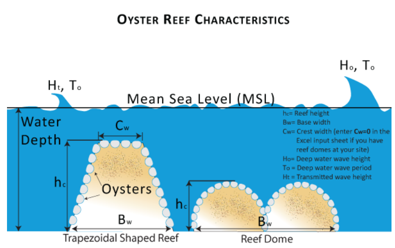


Fig. 5.24: Depiction of typical shapes of oyster reefs, along with definition of terms used in the input Excel sheet.

5.9.6 Running the Model

Setting up Workspace and Input Folders

These folders will hold all input and output data for the model. As with all folders for ArcGIS, these folder names must not contain any spaces or symbols. See the sample data for an example.

Note: The word *'path'* means to navigate or drill down into a folder structure using the Open Folder dialog window that is used to select GIS layers or Excel worksheets for model input data or parameters.

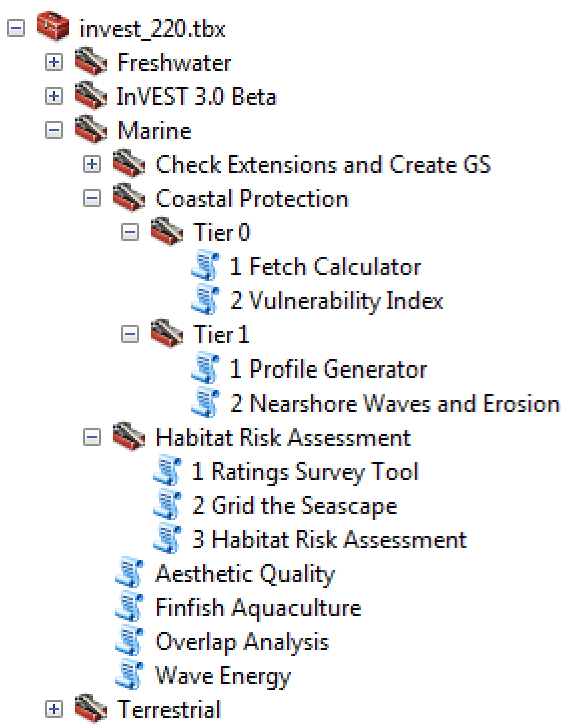
Exploring a Project Workspace and Input Data Folder

The */InVEST/CoastalProtection* folder holds the main working folder for the model and all other associated folders. Within the *CoastalProtection* folder there will be a subfolder named *'Input'*. This folder holds most of the GIS and tabular data needed to setup and run the model.

Creating a Run of the Model

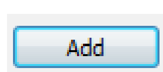
The following example of setting up the Erosion Protection model uses the sample data provided with the InVEST download. The instructions and screenshots refer to the sample data and folder structure supplied within the InVEST installation package. It is expected that you will have location-specific data to use in place of the sample data. These instructions provide only a guideline on how to specify to ArcGIS the various types of data needed and does not represent any site-specific model parameters. See the *Data Needs* section for a more complete description of the data specified below.

1. Click the plus symbol next to the InVEST toolbox.



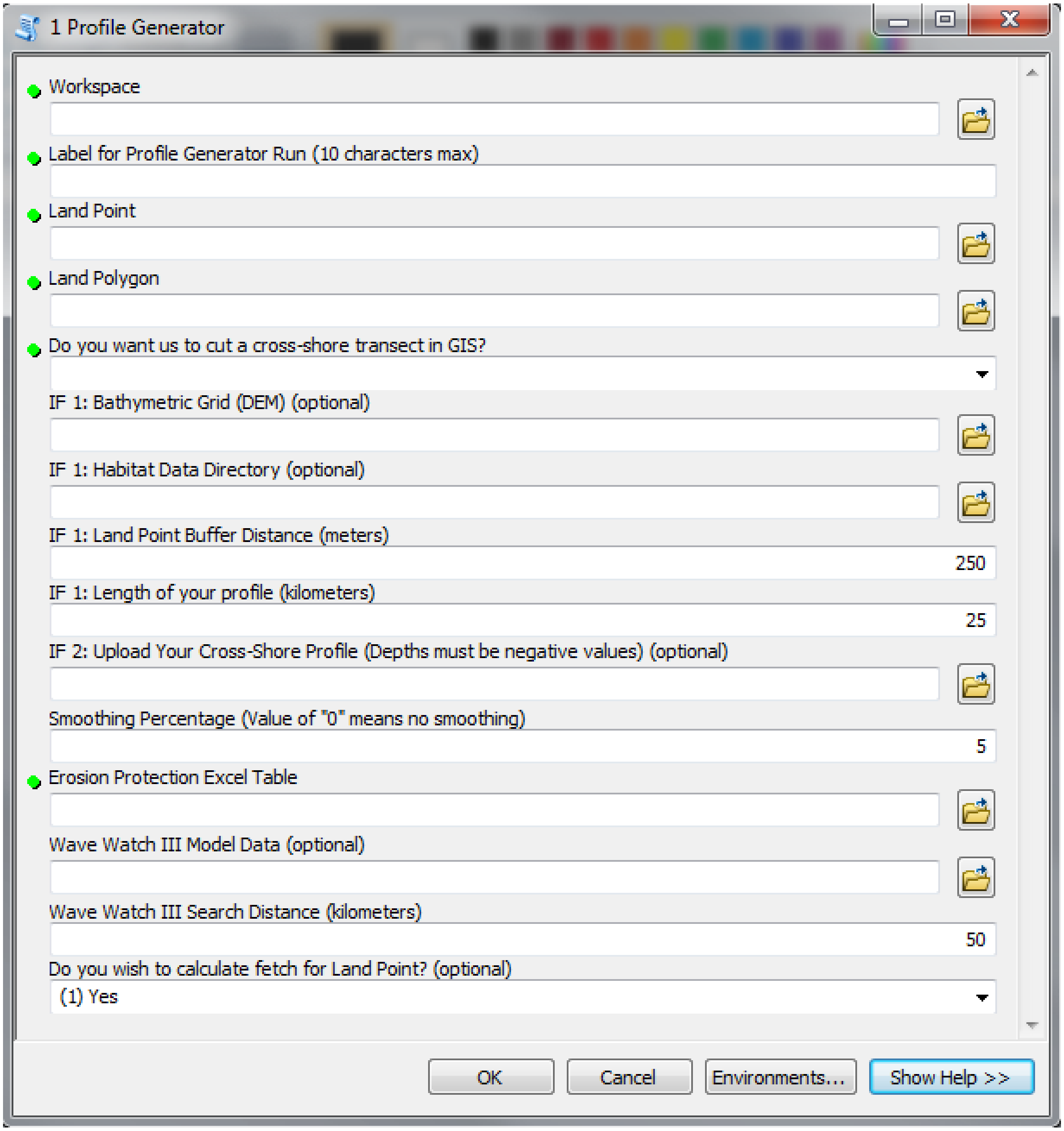
2. Expand the Marine, Coastal Protection, and toolsets. There are two scripts that you may want to run in succession: Profile Generator and Nearshore Waves and Erosion. Click on the Profile Generator script to open that model.

3. Specify the Workspace. Click on the Open Folder button  and path to the *InVEST/CoastalProtection/WCVI* folder. If you created your own workspace folder, then select it here.

   Click on the *WCVI* folder and click on 

 set the main model workspace. This is the folder in which you will find the "scratch" (intermediate) and "_ProfileGenerator_Outputs" (final outputs) folders after the model is run.

4. Specify the Label for Profile Generator Run. This string of text will be stripped of spaces and shortened to 10 characters. It will serve as the suffix to many of outputs. Type "Dune_2m" into the window.


1 Profile Generator
Workspace
Label for Profile Generator Run (10 characters max)
Land Point
Land Polygon
Do you want us to cut a cross-shore transect in GIS?
IF 1: Bathymetric Grid (DEM) (optional)
IF 1: Habitat Data Directory (optional)
IF 1: Land Point Buffer Distance (meters)
250
IF 1: Length of your profile (kilometers)
25
IF 2: Upload Your Cross-Shore Profile (Depths must be negative values) (optional)
Smoothing Percentage (Value of "0" means no smoothing)
5
Erosion Protection Excel Table
Wave Watch III Model Data (optional)
Wave Watch III Search Distance (kilometers)
50
Do you wish to calculate fetch for Land Point? (optional)
(1) Yes
OK
Cancel
Environments...
Show Help >>

5. Specify the Land Point. The model requires a land point shapefile to define the location for the analysis.

   Open the *InVEST/CoastalProtection/Input* data folder. Select the LandPoint_BarkSound.shp shapefile and click Add to make the selection.

6. Specify the Land Polygon. The model requires a land polygon shapefile to define the land and seascape for the analysis.

   Open the *InVEST/CoastalProtection/Input* data folder. Select the LandPolygon_WCVI.shp shapefile and click Add to make the selection.

7. Select '(1) Yes' that you wish to cut a cross-shore transect in GIS.

8. Specify the Bathymetric Digital Elevation Model (DEM) raster. The model requires a DEM raster file in order to cut a cross-shore transect in GIS. Click and path to the *InVEST/Base_Data/Marine/DEMs* data folder. Select the *claybark_dem* raster and click Add to make the selection.

9. Specify the Habitat Data Directory (optional). The model can use optional polygon shapefile that represents the location of various habitats. Click and path to the *InVEST/CoastalProtection/Input* data folder. Select the *NaturalHabitat* folder and click Add to make the selection.

10. Specify the Land Point Buffer Distance. The model requires this distance order to cut a perpendicular transect in GIS. The default distance is 250 meters, but may need to be modified depending on the site. You may change this value by entering a new value directly into the text box.

11. Specify the Length of your Profile (km). Provide the distance from your land point to a sufficiently deep adjacent location. If the location is sheltered by adjacent land masses, this length should be the distance, from the land point and orthogonal to the land polygon at that location, to the deepest point before crossing any land masses. This parameter defaults to 25 km but is site specific. You may change this value by entering a new value directly into the text box. For this example, 6 km is an appropriate length.

12. Specify the Smoothing Percentage. The model requires this value in order to smooth the bathymetry profile. The default percentage is 5, but may need to be modified depending on the DEM. You may change this value by entering a new value directly into the text box.

13. Specify the Erosion Protection Excel Table. The model requires you to specify information about your site for sediment size, tide elevation and habitats. A sample Erosion Protection Excel Table will be supplied for you.

    Click and path to the *InVEST/CoastalProtection/Input* data folder. Double left-click on the file *ErosionProtection_WCVI_BarkSound.xls*.

    Click Add to make the selection.

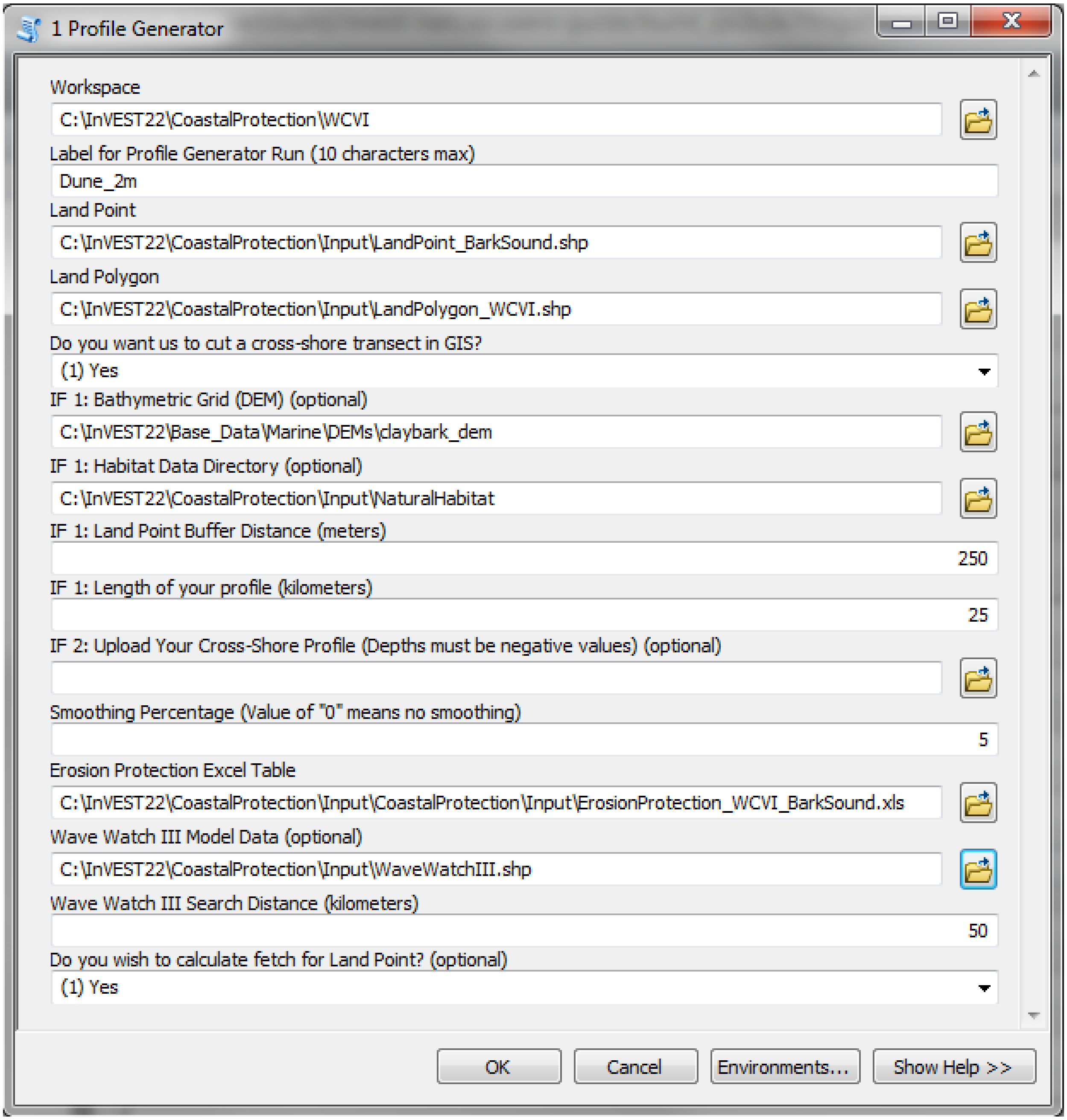
14. Specify the WaveWatchIII Model Data shapefile (optional). The model can use optional wind and wave statistics to represent oceanic conditions at a particular site. Click and path to the *InVEST/CoastalProtection/Input* data folder. Select the *WaveWatchIII.shp* shapefile and click Add to make the selection.

15. Specify the WaveWatchIII Search Distance. The model requires this search distance in order to find the closest WW3 point. The default distance is 50 km, but may need to be modified depending on the distance of your Land Point to the nearest WW3 point. You may change this value by entering a new value directly into the text box.

16. Select '(1) Yes' that you wish to calculate fetch for Land Point.

17. At this point the Profile Generator model dialog box is complete and ready to run.

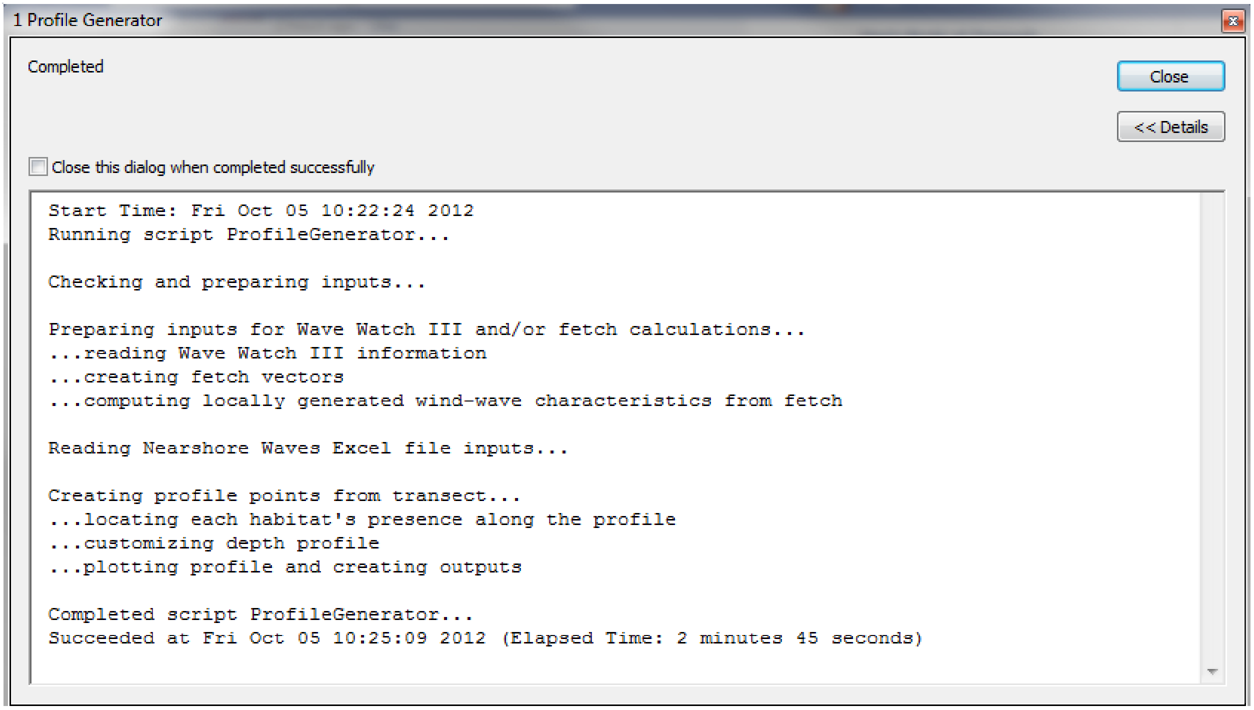
Click to start the model run. The Profile Generator will begin to run and a show a progress window with progress information about each step in the analysis. Once the model finishes, the progress window will show all the completed steps and the amount of time that has elapsed during the model run.



18. Now that your cross-shore profile has been created, you can click on the Nearshore Waves and Erosion script to open that model.

19. Specify the Workspace. Click on the Open Folder button and path to the *InVEST/CoastalProtection/WCVI* folder. If you created your own workspace folder, then select it here.

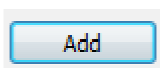
    Click on the *WCVI* folder and click on Add set the main model workspace. This is the folder in which you will find the "_WaveModel_Outputs" (final outputs) folders after the model is run.



20. Specify the Label for Nearshore Waves and Erosion run. This string of text will be stripped of spaces and shortened to 10 characters. It will serve as the suffix to many of outputs. Type “Dune_2m” into the window.

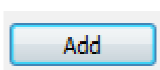
21. Provide the Erosion Protection Excel Table. The model requires you to specify information about site information and habitat management actions. A sample Erosion Protection Excel Table will be supplied for you.

    Click  and path to the *InVEST/CoastalProtection/Input* data folder. Double left-click on the file *ErosionProtection_WCVI_BarkSound.xls*.

    Click  to make the selection.

22. Specify a Cross-Shore Profile. The model requires a text file of a smoothed bathymetric and topographic transect. This can either be an output from the Profile Generator or a profile of your own.

    Click  and path to the *InVEST/CoastalProtection/Input* data folder. Double left-click on the file *InVEST/CoastalProtection/WCVI/_ProfileGenerator_Outputs/Dune_2m/html_txt/CreatedProfile_Dune_2m.txt*.

    Click  to make the selection.

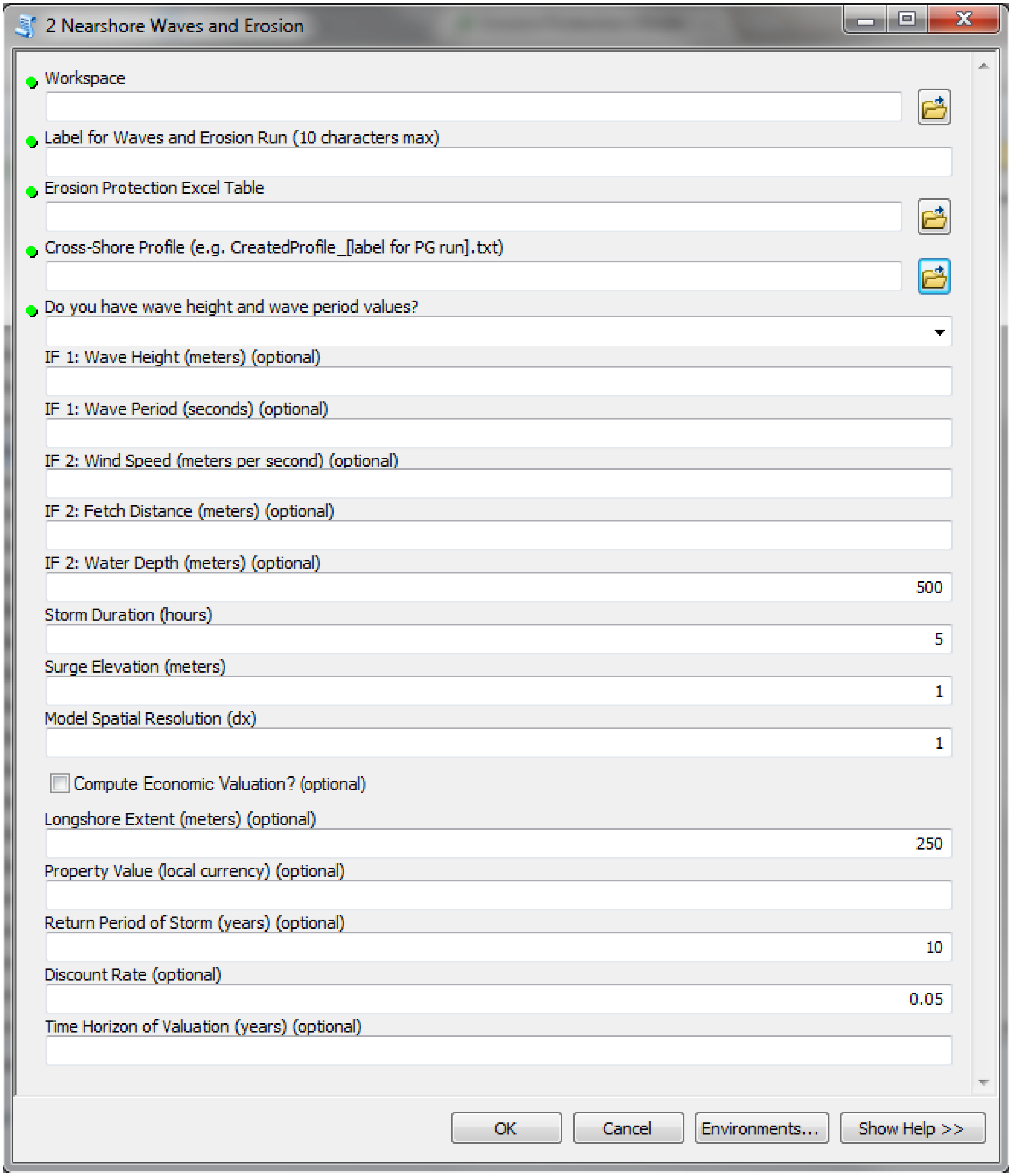
23. Select ‘(1) Yes, I have these values’ in answer to the question about whether you have wave height and period values.

24. Specify a Wave Height. Enter the wave height you wish to model. For this example, enter a value of “5” for this input.

25. Specify a Wave Period. Enter the wave period you wish to model. For this example, enter a value of “10” for this input.

    If ‘(2)No, Please compute these values from wind speed and fetch distance’ had been selected, which is only appropriate for sheltered shorelines, you would have to provide a wind speed, a fetch length (distance from point of interest to adjacent land masses), and an average depth in along the transect of interest.

26. Specify a Storm Duration. Please provide the duration of the storm (strong waves and surge) you are modeling. The default value is 5 hours but you can change this value by typing directly into the text box.


2 Nearshore Waves and Erosion
Workspace
Label for Waves and Erosion Run (10 characters max)
Erosion Protection Excel Table
Cross-Shore Profile (e.g. CreatedProfile_[label for PG run].txt)
Do you have wave height and wave period values?
IF 1: Wave Height (meters) (optional)
IF 1: Wave Period (seconds) (optional)
IF 2: Wind Speed (meters per second) (optional)
IF 2: Fetch Distance (meters) (optional)
IF 2: Water Depth (meters) (optional)
500
Storm Duration (hours)
5
Surge Elevation (meters)
1
Model Spatial Resolution (dx)
1
Compute Economic Valuation? (optional)
Longshore Extent (meters) (optional)
250
Property Value (local currency) (optional)
Return Period of Storm (years) (optional)
10
Discount Rate (optional)
0.05
Time Horizon of Valuation (years) (optional)
OK
Cancel
Environments...
Show Help >>

27. Specify the Surge Elevation. The model requires the elevation of the peak surge **relative to Mean Sea Level**. The default value is 1 meter but you can change this value to better represent the storm conditions and tidal range at your site. You can change the value by typing directly into the text box. For this example, enter a value of “3” for this input.

28. Specify the Model Spatial Resolution. The default resolution is 1m. If you would like a coarser resolution to improve run time, you can increase this value by typing a larger value into the text box.

29. Compute Economic Valuation? Check this box if you would like to approximate the change in erosion damages due to a change in habitat from a management action. This requires the remainder of the fields to be populated. If economic valuation is not desired, do not check this box.

30. Specify the Longshore Extent. If you wish to compute economic valuation, you will have to provide a distance along the shore where habitat, topo/bathy, forcing, habitat management actions, and property value are essentially uniform. This is a site specific parameter but 250m is the default value.

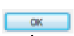
31. Specify the Property Value. If you wish to compute economic valuation, you will have to provide the property value of the nearshore land in your local currency per square meters. For this example enter a value of “12”.

32. Specify the Return Period of Storm. If you wish to compute economic valuation, you will have to provide the return period of the storm you are modeling (waves and surge). For example, if you are modeling surge and waves associated with the ‘100-year storm,’ enter a value of 100 here. For this example, enter a value of “25”.

33. Specify the Discount Rate. If you wish to compute economic valuation, you will have to provide a discount rate. The default value is 5% (0.05) but you are free to change this parameter if a different discount rate is more appropriate.

34. Specify the Time Horizon of Valuation. If you wish to compute economic valuation, you will have to provide the number of years into the future you would like to value the protective services of your habitat.

35. At this point the model dialog box is completed for a full run of the Nearshore Waves and Erosion portion of the Erosion Protection model.

    Click  to start the model run. The model will begin to run and a show a progress window with progress information about each step in the analysis. Once the model finishes, the progress window will show all the completed steps and the amount of time that has elapsed during the model run.

Viewing Output from the Model

Upon successful completion of the model, two new folders called “_ProfileGenerator_Outputs” and “_WaveModel_Outputs” will be created in each of the sub-models (Profile Generator and Nearshore Waves and Erosion) workspaces. They both contain a link to an html page that shows results of your run as well as various files that supplement the information on that html page. Output files are described in more detail in the cp-interpreting-results section.

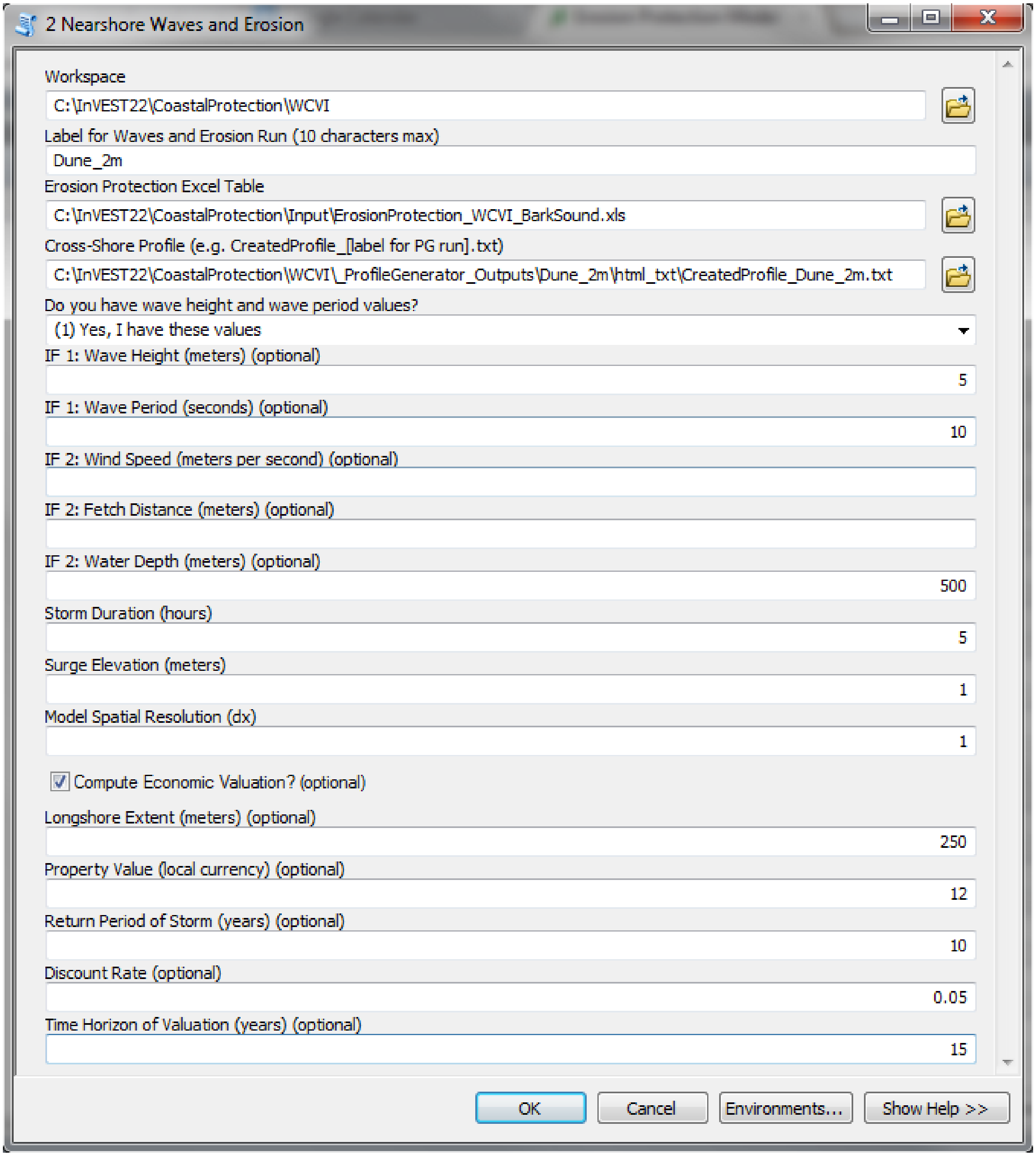
5.9.7 Interpreting Results

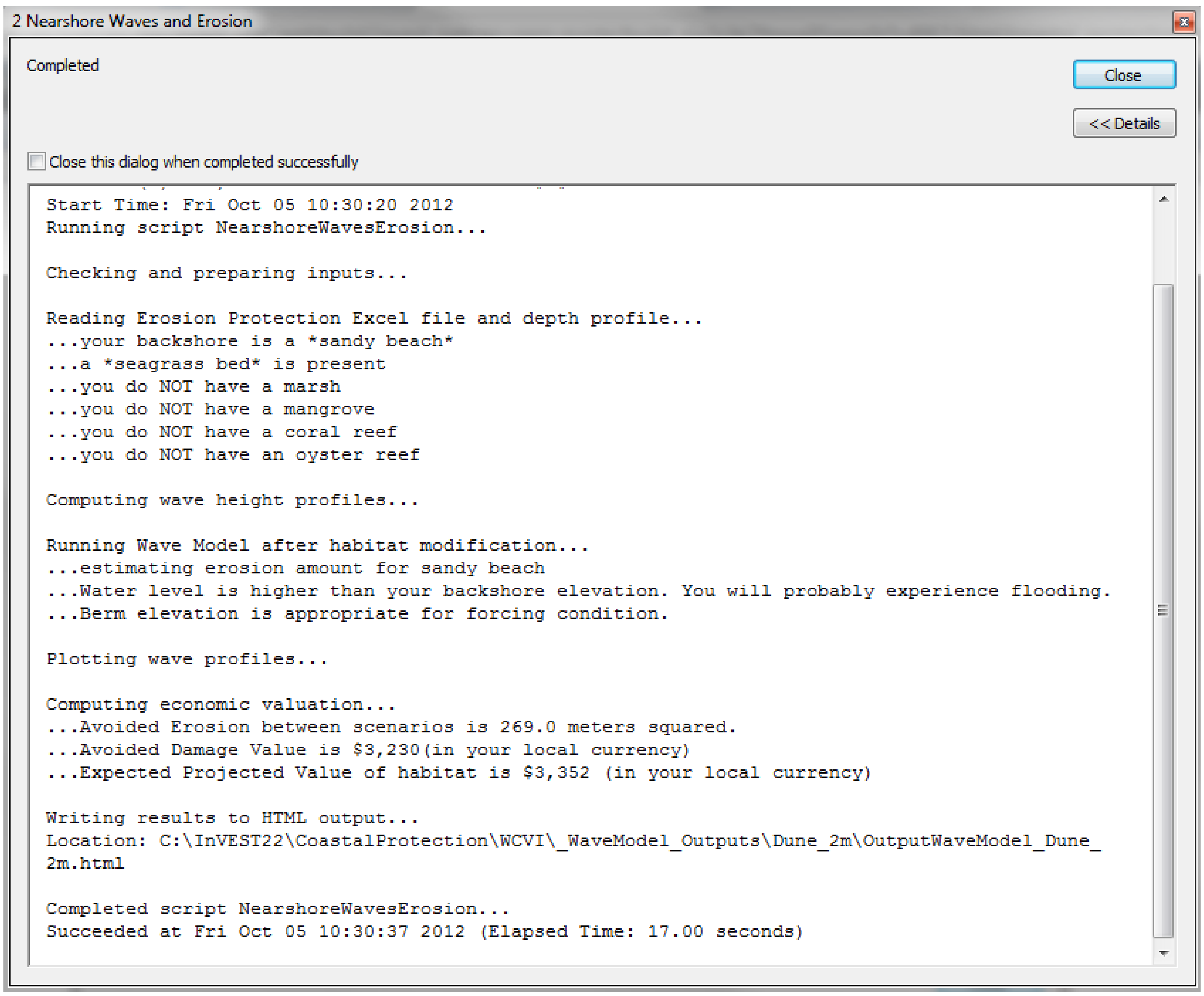
Model Outputs

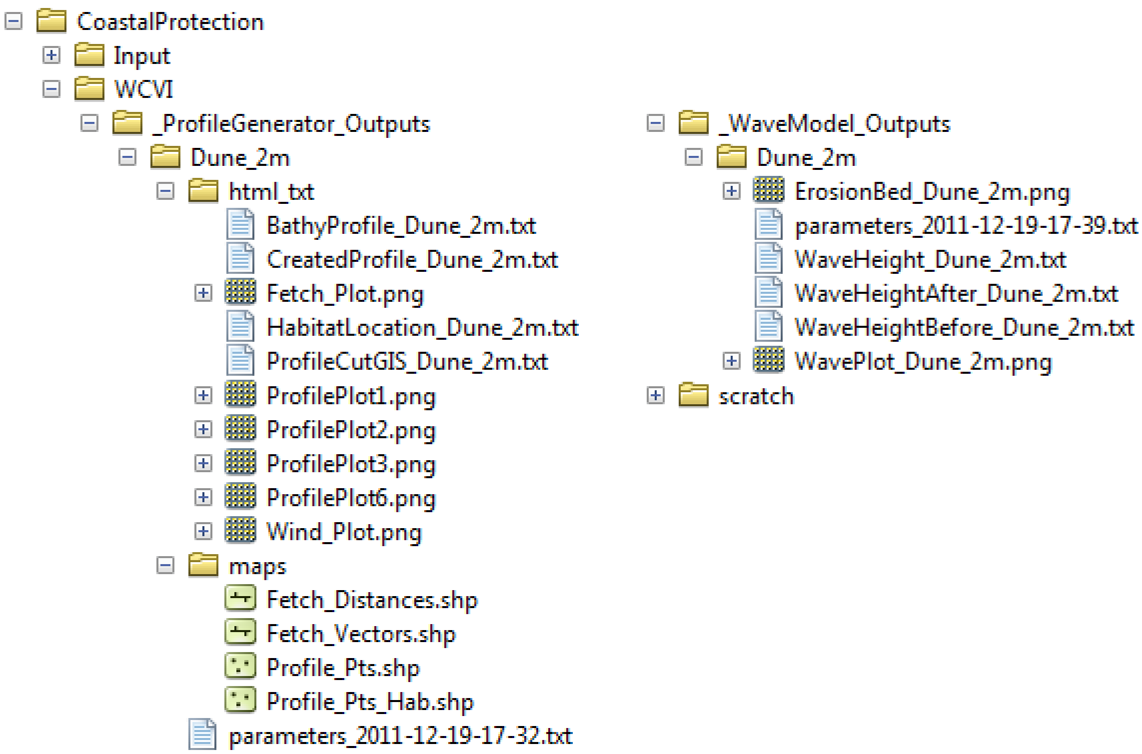
The following is a short description of each of the outputs from the Erosion Protection model. Each of these output files is saved in the output workspace directory you specified:

_ProfileGenerator_Outputs

This folder contains a sub-folder whose name is the “suffix label” you specified in this model’s interface. It contains two sub-folders: html_txt and maps.


2 Nearshore Waves and Erosion
Workspace
C:\InVEST22\CoastalProtection\WCVI
Label for Waves and Erosion Run (10 characters max)
Dune_2m
Erosion Protection Excel Table
C:\InVEST22\CoastalProtection\Input\ErosionProtection_WCVI_BarkSound.xls
Cross-Shore Profile (e.g. CreatedProfile_[label for PG run].txt)
C:\InVEST22\CoastalProtection\WCVI_ProfileGenerator_Outputs\Dune_2m\html_txt\CreatedProfile_Dune_2m.txt
Do you have wave height and wave period values?
(1) Yes, I have these values
IF 1: Wave Height (meters) (optional)
5
IF 1: Wave Period (seconds) (optional)
10
IF 2: Wind Speed (meters per second) (optional)
IF 2: Fetch Distance (meters) (optional)
IF 2: Water Depth (meters) (optional)
500
Storm Duration (hours)
5
Surge Elevation (meters)
1
Model Spatial Resolution (dx)
1
Compute Economic Valuation? (optional)
Longshore Extent (meters) (optional)
250
Property Value (local currency) (optional)
12
Return Period of Storm (years) (optional)
10
Discount Rate (optional)
0.05
Time Horizon of Valuation (years) (optional)
15
OK
Cancel
Environments...
Show Help >>


2 Nearshore Waves and Erosion
Completed
Close
<< Details
Close this dialog when completed successfully
Start Time: Fri Oct 05 10:30:20 2012
Running script NearshoreWavesErosion...
Checking and preparing inputs...
Reading Erosion Protection Excel file and depth profile...
...your backshore is a *sandy beach*
...a *seagrass bed* is present
...you do NOT have a marsh
...you do NOT have a mangrove
...you do NOT have a coral reef
...you do NOT have an oyster reef
Computing wave height profiles...
Running Wave Model after habitat modification...
...estimating erosion amount for sandy beach
...Water level is higher than your backshore elevation. You will probably experience flooding.
...Berm elevation is appropriate for forcing condition.
Plotting wave profiles...
Computing economic valuation...
...Avoided Erosion between scenarios is 269.0 meters squared.
...Avoided Damage Value is $3,230(in your local currency)
...Expected Projected Value of habitat is $3,352 (in your local currency)
Writing results to HTML output...
Location: C:\InVEST22\CoastalProtection\WCVI_WaveModel_Outputs\Dune_2m\OutputWaveModel_Dune_
2m.html
Completed script NearshoreWavesErosion...
Succeeded at Fri Oct 05 10:30:37 2012 (Elapsed Time: 17.00 seconds)



html_txt This folder contains two webpage links, figures used in the webpages, and three text files. + profile.html: This HTML file contains information summarizing the location of your site, as well as the information you entered in the model's interface and Excel input file. This output also contains figures showing the bathymetry profile created and/or smoothed by the Profile Generator Model, with close ups of the backshore area, when applicable. Also, if you have uploaded a folder of natural habitats *and** used the Profile Generator Model to cut a cross-shore transect for you from a DEM file, a table and figure are presented that indicate the X-coordinates of the beginning and end of where each natural habitat exists along the transect.

- fetchwindwave.html: This HTML file contains figures showing wind and fetch roses. It also contains information on fetch distances computed by the model, if you chose this option. There are also tables showing the average values of the maximum, as well as the top 10% and 25% wind speed and wave height extracted from the WW3 gage point closest to your site, if you uploaded that file. Finally, if you had the model compute fetch distances for you and uploaded WW3 data, this page also contains estimates of wind-generated wave height for each of the 16 equidistant sectors that make a full compass circle.
- FetchDistances_[suffix].txt: This text file contains information on fetch distances computed by the model. It has two columns. The first column shows the 16 directional sectors angles, and the second column has fetch distances associated with these sectors.
- BathyProfile_[suffix].txt: This text file is the smoothed bathymetric profile produced by the Profile Generator. It only contains values of water depths *below* MLLW (or the vertical datum of your bathy or topo/bathy DEM). The first column consists of X-values with X=0 at the shoreline, and the second column corresponds to depths values at the various cross-shore X distances.
- CreateProfile_[suffix].txt: This text file is the smoothed bathymetric and topographic profile produced by the Profile Generator. It differs from "BathyProfile_label.txt" because it has the backshore information. This backshore information was either provided by information in the Erosion Protection Excel Table or extracted from your DEM if you provided a seamless DEM. **We recommend that you use this profile as input in the Erosion Protection model.**
- ProfileCutGIS_[suffix].txt: This text file is the un-smoothed and un-processed raw profile that was cut by the

model, if you chose that option, before smoothing and/or the addition of backshore information. This file is useful if you want to see the quality of the GIS DEM data that you uploaded. If you have a good quality DEM layer that contains a high resolution representation of your area, this text file can also be useful and input in the wave model, as long as it is smoothed.

Maps

- Fetch_Vectors.shp: This polyline shapefile depicts the remaining fetch radials found in the seascape after being intersected with the Land Polygon input (landscape) you provided . The GIS starts with 144 vectors in total, at 2.5 degree increments, and erases all radials that overlap with the landscape.
- Fetch_Distances.shp: This polyline shapefile summarizes fetch distances for Land Point input over 16 directions that you specified.
- Profile_Pts.shp: This point shapefile represents the cross-shore transect that was cut by the GIS. Its attribute table contains depth information from both the raw and smoothed profiles.
- Profile_Pts_Hab.shp: This point shapefile represents the cross-shore transect that was cut by the GIS and then intersected with the habitat layers you provided. In the attribute table, columns for each of the six possible habitats are included. A value of "1" means a particular habitat is present at a point along the transect, while a "0" means it is not found.

_WaveModel_Outputs

This folder contains two useful outputs from the Nearshore Waves and Erosion model:

- OutputWaveModel_[suffix].html: This HTML file summarizes the information you entered as input in the model, including wave forcing and habitat management actions, and describes the outputs. It contains a figure depicting profiles of wave height (before and after habitat management action), as well as percent of wave attenuation and the location of your natural habitats along your bathymetry. It also provides a figure showing a profile of erosion or hourly rate of bed scour in your backshore area before and after management action. If valuation was selected, a table summarizing the value of your natural habitats before and after management action is presented.
- WaveHeight_[suffix].txt: This text file contains three columns showing distance from the shoreline and profiles of wave height over your bathymetry profile, before (second column) and after (third column) your management action.
- WaveHeightAfter_[suffix].txt: This text file contains two columns showing distance from the shoreline and profiles of wave height over your bathymetry profile, before after your management action.
- WaveHeightBefore_[suffix].txt: This text file contains two columns showing distance from the shoreline and profiles of wave height over your bathymetry profile, before your management action.

Parameter Log

Each time the module is run a text file will appear in the workspace folder. The file will list the parameter values for that run and be named according to the service and the date and time.

5.9.8 References

Armono, and K. Hall. (2003). Laboratory study of wave transmission on artificial reefs. Proc. Canadian Coastal Eng. Conf., Kingston, Canada

Apotsos, A., B. Raubenheimer, S. Elgar, and R.T. Guza. (2008). Testing and calibrating parametric wave transformation models on natural beaches, Coast. Eng., 55.

Alsina, J. M. and T. E. Baldock (2007). "Improved representation of breaking wave energy dissipation in parametric wave transformation models." Coastal Eng. 54(10).

Battjes, J.A. and M.J.F. Stive (1985). Calibration and verification of a dissipation model for random breaking waves, J. Geophys. Res., 90(C5).

Bradley, K., and C. Houser (2009), Relative velocity of seagrass blades: Implications for wave attenuation in low-energy environments, J. Geophys. Res., 114.

Burger, B. (2005). Wave Attenuation in Mangrove Forests, A Master's thesis publication, Delft U. of Technology, Civil Engineering and Geosciences. Cannon, M.G., J.M. Phelan, and M.A. Passaro. (1995) Procedural guidelines for estimating residential and business structure value for use in flood damage estimations. IWR Report 95-R-9, USACE. Water Resources Support Center, Alexandria, Virginia

Dean, R. G. and C. J. Bender (2006) "Static Wave Setup With Emphasis on Damping Effects by Vegetation and Bottom Friction", Coastal Eng. 13.

Dean, R.G. and R.A. Dalrymple. (2002) Coastal Processes with Engineering Applications. Cambridge University Press. 475pp.

de Vos, J.W. (2004). Wave attenuation in mangrove wetlands: Red River Delta, Vietnam. MSc thesis. Delft University of Technology

Gourlay, M.R. (1996a).Wave set-up on coral reefs. 1. Set-up and wave-generated flow on an idealised two dimensional reef. J. Coastal Eng. 27.

Gourlay, M.R. (1996b).Wave set-up on coral reefs. 2.Wave set-up on reefs with various profiles. J. Coastal Eng. 28.

Gourlay, M.R. (1997). Wave set-up on coral reefs: some practical applications. Proc. Pacific Coasts and Ports, Christchurch, 2, 959–964.

Keddy, P. A. (1982). Quantifying within-lake gradients of wave energy: Interrelationships of wave energy, substrate particle size, and shoreline plants in Axe Lake, Ontario. Aquatic Botany 14, 41-58.

Kobayashi, N., A. W. Raichle, and T. Asano (1993), Wave attenuation by vegetation, J. Waterw. Port, Coastal Ocean Eng., 119

Koch, E.W., L.P. Sanford, S.N. Chen, D.J. Shafer and J.M. Smith (2006). Waves in seagrass systems: review and technical recommendations. US Army Corps of Engineers Technical Report. Engineer Research and Development Center, ERDC TR-06-15, 82 p.

Komar, P.D. (1998) Beach Processes and Sedimentation, Prentice Hall, Upper Saddle River, N.J., 543pp.

Kriebel, D. L., and R.G. Dean. (1993). Convolution method for time dependent beach-profile response. J. Waterw., Port, Coastal, Ocean Eng., 119(2)

Mazda, Y., E. Wolanski, B. King, A. Sase, D. Ohtsuka, and M. Magi (1997). Drag force due to vegetation in mangrove swamps. Mangroves and Salt Marshes 1:193–99.

McLachlan, A. and Dorvlo, A. (2005). Global patterns in sandy beach macrobenthic communities. Journal of Coastal Research 21, 674-687.

Mendez, F. J., and I. J. Losada (2004), An empirical model to estimate the propagation of random breaking and nonbreaking waves over vegetation fields, Coastal Eng., 51

Narayan, S. (2009). The Effectiveness of Mangroves in Attenuating Cyclone – induced Waves, Master's thesis, Delft U. of Tech., Civil Eng. and Geosciences

Stockdon, H.F., R.A. Holman, P.A. Howd, and A.H. Sallenger. (2006). Empirical parameterization of setup, swash, and runup. Coastal Engineering, 53

Thornton, E. and Guza, R.T. (1983). Transformation of Wave Height Distribution. Journal of Geophysical Research 88(C10)

Tolman, H.L. (2009). User manual and system documentation of WAVEWATCH III version 3.14, Technical Note, U. S. Department of Commerce Nat. Oceanic and Atmosph. Admin., Nat. Weather Service, Nat. Centers for Environmental Pred., Camp Springs, MD.

U.S. Army Corps of Engineers (USACE) (2002). U.S. Army Corps of Engineers Coastal Engineering Manual (CEM) EM 1110-2-1100 Vicksburg, Mississippi.

Van der Meer, J.W., Briganti, R., Zanuttigh, B. and Wang, B. (2005). Wave transmission and reflection at low crested structures: design formulae, oblique wave attack and spectral change, Coast. Eng., 52.

Whitehouse, R., Soulsby, R.R., Roberts, W., Mitchener, H. (2000). Dynamics of Estuarine Muds. H. R. Wallingford, UK Wiegel, R.L. 1964 Oceaographical Engineering, Prentice-Hall, Englewood Cliffs, NJ.

5.9.9 Appendix A

Beach Survey with "Emery Boards"

(Adapted from *Beach Profiling with "Emery Boards" and Measuring Sand Grain Size*, 2005, Florida Center for Instructional Technology, University of South Florida)

The simplest technique to measuring a beach profile is known as the **"Emery board"** method, developed by a famous coastal scientist named K.O. Emery. As depicted in Figure 1 the apparatus consists of two stakes connected by a rope of known length (5m or 10m). This length sets the measurement interval for individual data points along the profile. Each stake has a measurement scale which runs from 0 at the top, down to the bottom of the stake. It is recommended to use Metric units. This approach may seem simple, but it provides reasonably accurate measurements of beach profiles. It also has the advantages of light, inexpensive, equipment, which can be easily carried to distant survey sites, for very rapid surveys.

The technique of measuring sand size will be conducted in the field with the use of sand gauge charts. These are small, credit-card sized, plastic charts with calibrated samples of sieved sand mounted on the face. By using a hand-lens and sand gauge chart, it is possible to compare samples from the beach with calibrated samples on the chart for an estimate of size range. Sand gauge charts are available from a number of vendors. One such distributor is ASC Scientific.

Materials

To build a set of "Emery boards", all that is needed are two pieces of wood of equal length and a rope of known length. (Boards slightly smaller than observers will work well (~1.6m).) Tie a loop in each end of the rope, which can easily slide up and down the two boards. Measuring down from the top of each board, use a marker and a ruler to draw and label the "graduations" (marks of equal length). An appropriate graduation interval is every two centimeters. Additionally, one can attach a small level to the rope to help ensure it is horizontal (for example).

Method

At the very minimum, two people are necessary to conduct a survey, but three are preferable. Team members should separate themselves into a **"seaward surveyor"**, a **"landward surveyor"**, a **"geotechnical engineer"** and a **"data recorder"**. The "seaward surveyor" is responsible for holding the seaward board and ensuring that the rope is level between the two boards (by sliding the loop up or down) when fully extended. The "landward surveyor" is responsible for holding the landward board, sighting over the seaward board to the horizon, and shouting out the measurement (cm down from the top of the landward board) to the "data recorder". The "geotechnical engineer" is responsible for moving with the "seaward surveyor" to collect a sand sample, and identify it using the hand lens on the basis of its size comparison to the sand gage chart. The "data recorder" should keep organized notes of each measurement

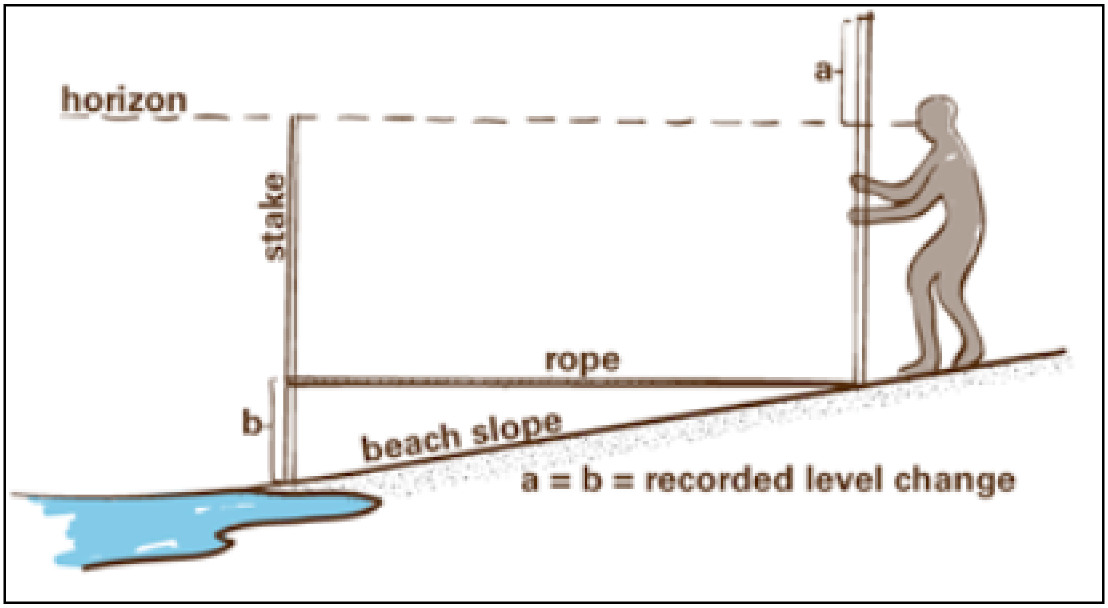


Fig. 5.25: Illustration of the Emery Board technique

including **horizontal distance (x), measurement of change in elevation (a), cumulative change in elevation of all measurements, and sand size at each location**.

Starting at the landward extent of the survey region (baseline), cross-shore data points of elevation and sand size are collected at the sampling interval determined by the length of the rope (distance between the two boards at full extension). Collect at least 5 cross shore data points. Collect more than 5 cross shore data points if the beach is wide. If the beach is sloping downward toward the sea, the observer sights across the top of the seaward board to the level of the horizon, and determines the distance **(A1)** from the top of the landward board to the sightline in the following figure (or distance **(a)** in Figure 1).

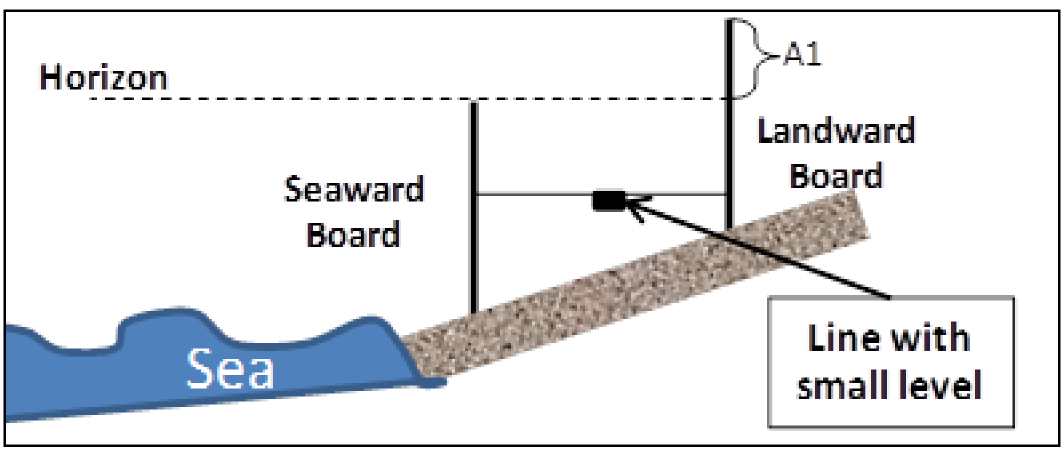


Fig. 5.26: Same as previous figure. Find distance A1 from top of board to eye such that eye, top of board 2 and horizon are aligned. Line must be horizontal.

If the beach is locally sloping upward in the offshore direction, then **(A2)** is measured on the seaward board and the sighting is with the horizon over the top of the landward board (next figure). If horizon cannot be found on landward side, then observer on landward aligns his/her eye with pointer (pen or other thin sharp object) adjusted and held by observer on seaward side and horizon to form a horizontal line. Observer on seaward side then reads distance A2, which should be recorded as negative to indicate upward slope.

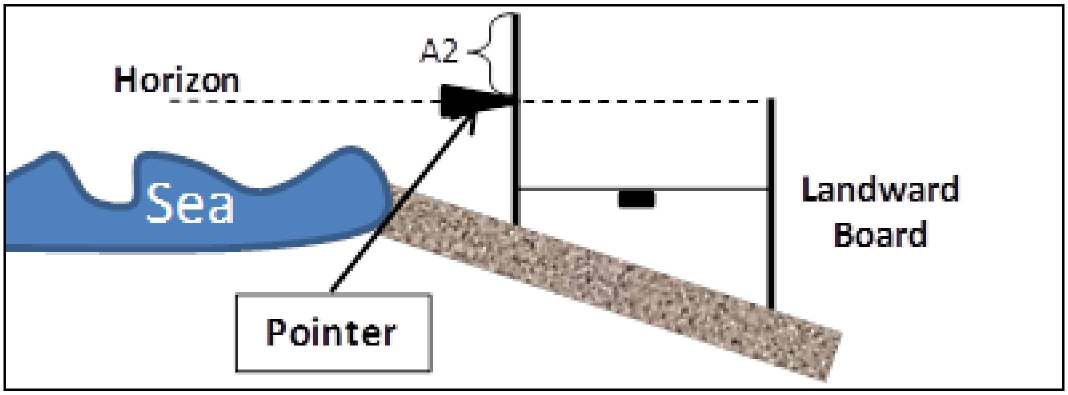


Fig. 5.27: Same as two previous figures. Find distance A2 from top of board to pointer such that eye at top of board 1, pointer and horizon are aligned. Line must be horizontal.

In addition, the "data recorder" should make notes of the time of observations and such things such as presence/absence and type of beach debris (kelp, wood etc.). Also, the "data recorder" should take note of the maximum landward extent of these debris if they were freshly deposited, as an indication of position of high tide. High tide location can also be guessed by looking for position of wet/dry sand barrier. If the team has a portable GPS unit, the "data recorder" should note the coordinate of this high water mark, or if there are repeated measurements at the same site, the "data recorder" should evaluate its distance from known landmark. Finally, the "data recorder" should make note of position (GPS or meters) of position of landward board during first measurement, of seaward board after last measurement, and position of water level.

Recording and Processing Data

Assuming that the rope is 10m long, an example log looks as follow, where positive values are A1 measurements (sloping down), and negative values are A2 measurements (sloping up):

Measurement	1	2	3	4	5	6
Value	5	-2	-1	3	8	10

Based on these values, a beach profile can be constructed by performing the following operations:

Measured values are in column 1, and cumulative distance between measurements is in Column 2 (assuming rope is 10m long). In Column 3 we estimate 1/Slope, using DX=length of rope=10m. For example, slope of 1st measurement is 1/2. In Column 4, we estimate beach profile, assuming that zero is located at point where first measurement is taken. In Column 5 we estimate beach profile again, assuming that zero is last point measured. This last column is used to plot profile of beach as function of X, as shown in the following figure.

Finally, if repeated measurements are made at the same time, it is recommended to continuously log time of measurement, and positions of board at beginning and end of measurement, as well as high water mark. These should be indicated on beach profile, if possible. Also, by looking at tide chart, it is possible to estimate high water level during period of measurement, and use this info to convert beach profile values accordingly.

Value DY	X Dist	1/Slope=DX/DY	Elevation1	Elevation2
5	10	2.0	-5	40
-2	20	-5.0	-3	42
-1	30	-10.0	-2	43
3	40	3.3	-5	40
8	50	1.3	-13	32
10	60	1.0	-23	22
7	70	1.4	-30	15
9	80	1.1	-39	6
6	90	1.7	-45	0

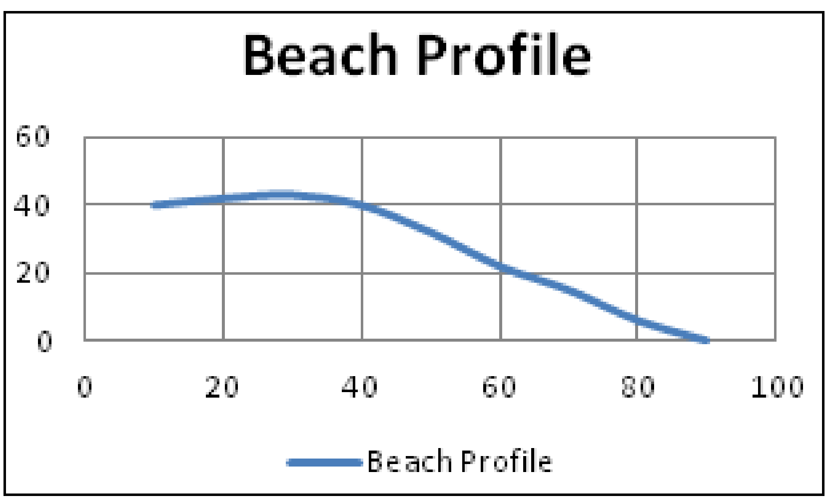


Fig. 5.28: Example beach profile measured with Emery Board. Zero is last point measured.

5.10 Managed Timber Production

5.10.1 Summary

An important ecosystem service provided by forests is the production of timber. This model analyzes the amount and volume of legally harvested timber from natural forests and managed plantations based on harvest level and cycle. The valuption model estimates the economic value of timber based on the market price, harvest and management costs and a discount rate. and calculates its economic value. Limitations of the model include assumptions that timber harvest production, frequency, prices, and costs are constant over time.

5.10.2 Introduction

Commercial timber production is a valuable commodity provided by forests, with the potential to generate significant revenue for those with legal rights to harvest. The scale and nature of timber production varies from large privately-operated single-species plantations to small community-managed harvests from natural forest that retains its ecological structure and function. Whether timber production occurs on a managed plantation or a natural forest, managing the intensity and rate of timber harvest is critical to sustaining this service, as well as the supply and value of other services provided by forests, such as water purification, carbon sequestration, and bush meat habitat. Maximizing profits requires information about the volume and species of wood removed in each harvest period, timber prices, and management costs.

5.10.3 The Model

The model is designed for cases where an entity (e.g., a government, a tribe, a community, a private timber company) has a formally recognized right to harvest *roundwood* from a forest. According to FAOSTAT (http://faostat.fao.org/), roundwood is wood in its natural state as felled, or otherwise harvested, with or without bark, round, split, roughly squared or in other forms. It comprises all wood obtained from removals. This model's output maps the net present values of forests' legally recognized harvests over some user-defined time interval. This model is very simple and designed for cases where little data on harvest practices and tree stand management exists. If you have access to detailed harvest and forest management data, you may want to use an alternative model.

Timber harvest by entities that do not have a formally recognized harvesting right is not accounted for in this model. This type of wood harvest, whether it is illegal or occurs in forest areas where property rights are either not defined or not well enforced, is dealt with in the Open Access Timber and Non-Timber Products Model (to be released soon).

How it Works

This model can be used in one of two ways. First, it can be used to model the expected value of a stream of harvests from a forest plantation over a user-defined time interval. A forest plantation is typically managed in such a way that merchantable or usable wood can be harvested at regular periods over an indefinite period. Three characteristics of a plantation forest are: 1) species mix has been reduced to a single or a few of the fastest growing species; 2) the oldest wood in the plantation is harvested and the rest of the wood is left to mature; 3) the areas of a plantation that have been clear-cut are replanted with the managed species soon after the clear-cut; and 4) a more or less even distribution of tree ages (e.g., if the oldest trees in the stand are 20 years old, a quarter of the stand is 1-5 year old, a quarter of the stand is 6-10 years old, a quarter of the stand is 11-15 years old, and a quarter of the stand is 16-20 years old).

Second, the InVEST Managed Timber Production Model can be used to calculate the expected value of timber harvests from primary, natural forests. By primary, natural forests we mean areas that, at least at the beginning of a harvest cycle, retain much of their natural structure and function. These could include forests that, at least at the beginning of a harvest cycle, are being used by local communities and tribes for small-scale timber and non-timber forest product harvest. In some cases these forests may become subject to large scale timber harvest because they are to transition to more managed forests (i.e., forest plantations as described above) or some other non-forest development that requires a clear-cut, such as agricultural or residential expansion. (This does not include forests that are slashed and burned, given that the felled and burned wood is not used to create a product). In other cases, concessions to clear-cut certain areas of a natural forest or selectively log a natural forest may be held by entities. In these cases an altered version of the natural forest would remain on the landscape into the future. Examples of this type of harvest include logging of rainforests in the Amazon or Malaysia for land conversion or in Indonesia to establish palm plantations, and selective clear-cutting of rainforests in Malaysia.

The model runs on a vector GIS dataset that maps parcels on the landscape that are, or are expected to be, used for legal timber harvest over a user-defined time period. These timber parcels can include a whole forest or just part of a forest. In any case a parcel should only include the portion of a forest that is formally designated, zoned, or managed for harvest. Each timber harvest parcel is described by its harvest levels (*Harv_mass* and *Perc_Harv* in the production table; see Data Needs section below), frequency of harvest (*Freq_harv*), and harvest and management (or maintenance) costs (*Harv_cost* and *Maint_cost*, respectively) (Fig. 1).

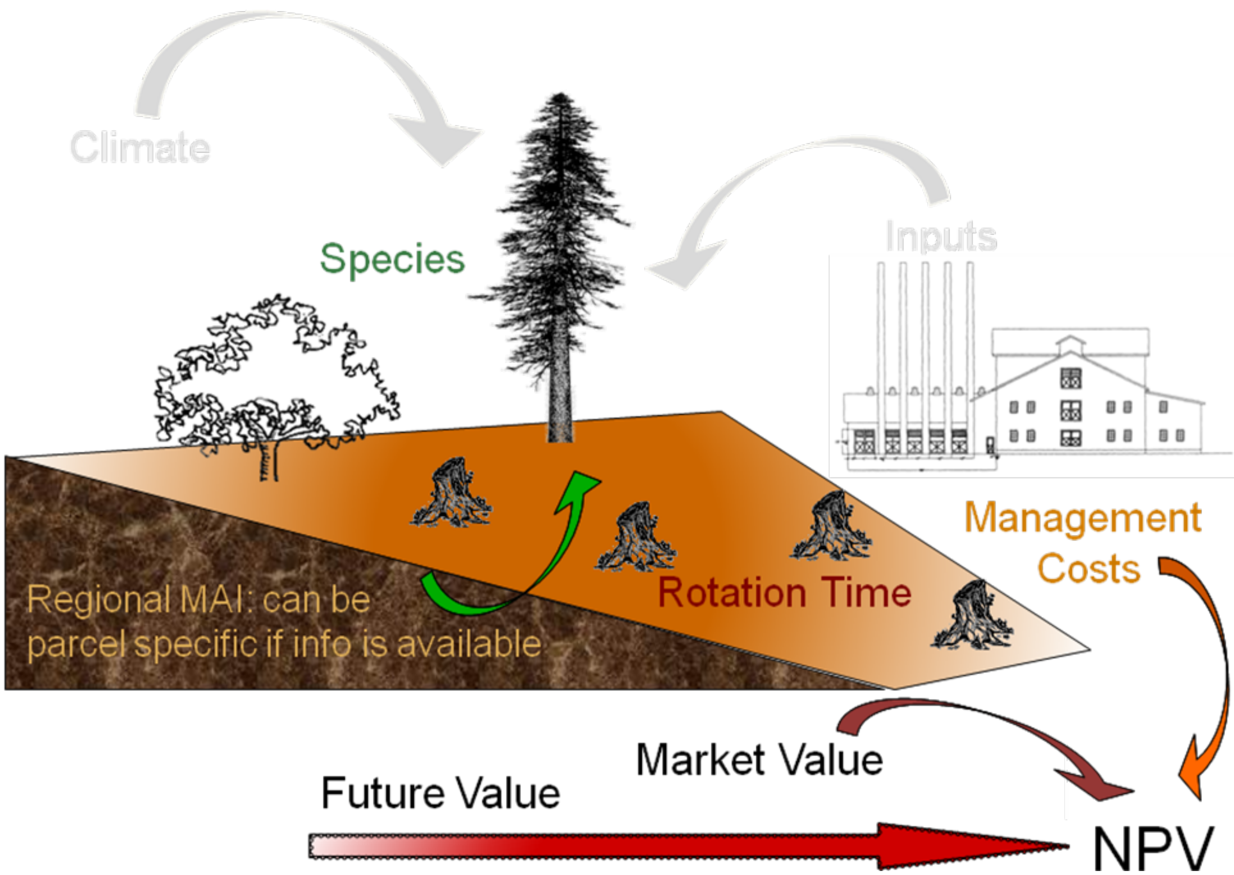


Fig. 5.29: *Figure 1. Conceptual diagram of the Managed Timber Production model. Parameters represented in color are included in the model, while those in gray are not.*

The timber parcel map can either be associated with a current (sometimes referred to as "base") L map is given by *yr_cur*) or with some future LULC scenario map (where the year associated with the future LULC map is given by *yr_fut*). If the timber parcel map is associated with the current LULC map the model calculates, for each timber parcel, the net present value (NPV) of harvests that occurred between the current year and some user-defined date, assuming that harvest practices and prices are static over the time interval modeled. If the timber parcel map is associated with a future scenario LULC map the model calculates, for each timber parcel, the NPV of harvests that occurred between the future date and some user-defined date, again assuming that harvest practices and prices do not change over the user-defined time interval. The model produces the NPV of harvests in the currency of either the current year or future year, depending on whether the user inputs a current or future LULC map. For example, if the selected year for the future scenario is 2050 and the dollar is the currency used to value timber harvests, then the NPV of harvests from 2050 to some user-defined later than 2050 is given in year 2050 dollars.

Limitations and Simplifications

This model assumes that the percent of the forest harvested each harvest period, the mass of timber harvested each harvest period, the frequency of each harvest period, and harvested related prices and costs remain constant in each timber parcel over the user-defined time period. In reality, each of these variables can change from year to year. For example, the mix of species harvested from a forest could change from one harvest period to the next and this could affect everything from the amount of wood harvested to the composite price received for the timber. In addition, un-modeled disturbances, such as forest fires or disease, or occasional managed thinning can have a major impact on harvest levels from a forest parcels.

Some of these limitations can be addressed by constraining the length of the time period used to assess harvests in parcels. For example, if the current year is 2000 and only the expected harvests until 2010 are valued, any unaccounted changes in timber harvest management or price changes may be minor. At this point a future 2010 LULC and timber parcel map could be evaluated with the timber model looking 10 years ahead again, from 2010 to 2020. The future timber parcel map could include any changes in timber management and prices that occurred between 2000 and 2010. This process could be repeated for successive decades until, for example, 2050. Successive model runs with decadal time intervals until 2050, and the ability to change harvesting behavior and prices, will better approximate harvesting practices on the landscape than just running the model once from 2000 to 2050.

Further, given the expected variation in harvest management practices and prices over the modeled time interval, it is suggested that the user use mean values for each model input. The mean is typically the best summary of the distribution of expected values for a variable. For example, if it is known that harvests from a timber parcel over time will involve various species it is possible to set the timber price for that parcel equal to the average expected price for all harvested species.

5.10.4 Data Needs

The model requires a GIS polygon file (a vector database) demarcating timber parcels. Unique timber parcels can be distinguished by differences in the percent of the parcel harvested each harvest period, the mass of wood removed each harvest period, the species of trees removed, or the costs of managing and harvesting wood from the parcel. These attributes, along with timber prices and the time interval for analysis, should be included as a separate CSV table.

1. **Timber parcels (required)**. A GIS dataset (vector) that indicates the different timber parcels on the landscape. Each parcel should be given a unique identifier.

   Name: file can be named anything

   File type: standard GIS polygon file (e.g., shapefile), with a unique identifier code for each polygon.

   Rows: each row is a timber parcel.

   Columns: Each parcel should be identified with a unique ID.

   Sample data set: \Invest\Timber\Input\plantation.shp

2. **Production table (required)**. A CSV table of information about the timber parcels on the landscape.

Name: file can be named anything

File type: *.CSV.

Rows: each row is a different parcel.

Columns: contain an attribute for each parcel and must be named as follows:

1. *Parcel_ID*: Same as timber parcel ID in #1. IDs must match the parcel IDs used in the polygon map. User must select this field as a model input.
2. *Parcl_area*: The area of the timber parcel in hectares.
3. *Perc_harv*: The proportion of the timber parcel area that is harvested each harvest period; units are integer percent.
4. *Harv_mass*: The mass of wood harvested per hectare (in metric tons (Mg) ha^{-1}) in each harvest period.
5. *Freq_harv*: The frequency of harvest periods, in years, for each parcel.
6. *Price*: The marketplace value of the wood harvested from the parcel ($^{-1}$). This price should reflect what is paid to the harvesters at mills or at other timber processing and collection sites. If a harvest includes multiple species, each with its own price, a weighted price should be used, where weights are given by the expected relative mix of the species in the harvest. Any value derived from pre-commercial thins should be included in Maint_cost (see below).
7. *Maint_cost*: The annualized cost ha^{-1} of maintaining the timber parcel, if any. Costs may include the periodic costs to replant, treat and thin the stand, plus the cost to harvest, treat slash, and deliver wood to a processing facility. Other costs may include taxes, pest treatments, etc. If commercial thins before the main harvest produce product that has market value, the annual ha^{-1} value of these harvests should be subtracted from Maint_cost. If the harvest comes from a natural forest that is not managed for timber production Maint_cost may be 0. (Actual stand maintenance costs may vary from year to year in a forest (e.g., in some years portions of a managed stand may have to be thinned prior to harvest and in other years anti-pest measures may have to be employed), an annualized value "smoothes" this temporal variation in maintenance costs.)
8. *Harv_cost*: The cost (ha^{-1}) incurred when harvesting Harv_mass.
9. *T*: The number of years from yr_cur or *yr_fut* that parcel harvests will be valued. If the parcel is in an even age rotation managed plantation, *T* can be any number, although as we explain below, we recommend against large *T*. If the harvest is expected to be an immediate one time clear cut *T* = 1. If a series of clear cuts in a natural forest are occurring or are expected, *T* can be no greater than the number of years that harvest of the natural stand can continue given Perc_harv and *Freq_harv*. For example, if a natural stand is going to be replanted as a single species plantation or allowed to regenerate naturally before being harvested again in the future, *T* for the harvest of the natural stand can be no larger than 7 if Perc_harv = 33.3 and *Freq_harv* = 3 (assuming a harvest takes place in years 1 (*yr_cur* or *yr_fut* depending on the associated LULC map), 4, and 7).
10. *Immed_harv*: This attribute answers whether a harvest occurs immediately – whether a harvest occurs in *yr_cur*, or whether the user is evaluating a forest parcel associated with a future LULC scenario occurring in *yr_fut*. Answer yes (entered as YES or Y) or no (entered as No or N) to whether a harvest should be calculated for *yr_cur* or *yr_fut*. If yes, then the NPV of harvest in the parcel includes a harvest in *yr_cur*, otherwise the first harvest accounted for in the parcel's NPV occurs *Freq_harv* years into the into time interval T.
11. *BCEF*: An expansion factor that translates the mass of harvested wood into volume of harvested wood. The expansion factor is measured in Mg of dry wood per m3 of wood. The expansion factor is a function of stand type and stand age (this factor is know as the biomass expansion factor in the literature). If you do not have data on this expansion factor you can use the $BCEF_R$ row in table 4.5 of IPCC (2006). Otherwise, set this expansion factor equal to 1 for each parcel.

Sample data set: \Invest\Timber\Input\plant_table.csv

3. **Market Discount Rate (optional – required for valuation)**. This number is not supplied in a table, but instead is input directly through a tool interface (Labeled "Market discount rate (%)" in the tool interface.) The market discount rate reflects society's preference for immediate benefits over future benefits (e.g., would you rather receive $10 today or $10 five years from now?). The tool's default value is 7% per year, which is one of the rates recommended by the U.S. government for evaluation of environmental projects (the other is 3%). However, this rate will differ depending on the country and landscape being evaluated. It can also be set to 0% if so desired.

To calculate NPV for a forest parcel a series of equation are used. First, we calculate the net value of a harvest during a harvest period in timber parcel *x*,

$$VH_x = \frac{Perc\_harv_x}{100}(Price_x \times Harv\_mass_x - Harv\_cost_x) \tag{5.67}$$

where VH_x is the monetary value (ha^{-1}) generated during a period of harvest in x, $Perc\_harv_x$ is the percentage of x that is harvested in each harvest period (converted to a fraction), Pricex is the market price of a Mg of timber extracted from x, $Harv\_mass_x$ is the Mg ha^{-1} of wood removed from parcel x during a harvest period, and Harv_costx is the cost (ha^{-1}) of removing and delivering $Harv\_mass_x$ to a processing facility or transaction point. In general, $Harv\_mass_x$ will be given by the aboveground biomass (Mg ha^{-1}) content of the forest stand less any portion of the stand that is left as waste (e.g., stems, small braches, bark, etc.). For example, assume a company plans to clear-cut 10% of a native forest block in each harvest period, Pricex is expected to be $10 $^{-1}$, $Harv\_mass_x$ is 800 Mg ha^{-1}, and Harv_costx = $5,000 ha^{-1}. The net value created during a harvest period is given by,

$$VH_x = 0.1 \times (10 \times 800 - 5000) = 300 \tag{5.68}$$

A harvest period is a sustained period of harvest followed by a break in extraction. Plantation forests tend to have a harvest period every year. More natural forests may have more intermittent periods of harvest (e.g., a pulse of harvest activity every 3 years). The periodicity of harvest periods in parcel x is given by the variable $Freq\_harv_x$.

The variable $Freq\_harv_x$ is used to convert the per hectare value of the parcel (math:*VH_x*) into a stream of net harvest revenues, which is then aggregated and discounted appropriately. Specifically, the NPV (ha^{-1}) of a stream of harvests that engender math:*VH_x* intermittingly from *yr_cur* or *yr_fut* to T_x years after *yr_cur* or *yr_fut* is given by:

$$NPV_x = \sum_{s=0}^{ru\left(\frac{T_x}{Freq\_harv_x}\right)-1} \frac{VH_x}{\left(1+\frac{r}{100}\right)^{Freq\_harv_x \times s}} - sum_{t=0}^{T_x-1}\left(\frac{Mait\_cost_x}{\left(1+\frac{r}{100}\right)^t}\right) \tag{5.69}$$

where "ru" means any fraction produced by T_x / $Freq\_harv_x$ is rounded up to the next integer, $Freq\_harv_x$ is the frequency (in years) of harvest periods, *r* is the market discount rate, and Maint_costx is the annualized cost (ha^{-1}) of managing parcel *x*. Continuing our earlier example, where math:*VH_x* = 300, if we set $Freq\_harv_x$ = 1 (a harvest period occurs every year), T_x equal to 10 (T_x can be no larger than 10 because the native forest will be completely gone in 10 years given $Perc\_harv_x$ = 10%), r equal to 7%, and Maint_costx equal to $50 ha^{-1}, then the NPV of the stream of math:*VH_x* is,

$$NPV_x = \sum_{s=0}^{9}\frac{300}{1.07^s} - \sum_{t=0}^{9}\frac{50}{1.07^t} \tag{5.70}$$

On the other hand, assume $Freq\_harv_x$ = 3 (a 10% harvest of the timber parcel occurs every 3 years) and all other variables are as before, then,

$$NPV_x = \sum_{s=0}^{ru\left(\frac{10}{3}\right)-1}\frac{300}{1.07^{3\times s}} - \sum_{t=0}^{9}\frac{50}{1.07^t} \tag{5.71}$$

In other words, a harvest period occurs in years 1 (*yr_cur* or *yr_fut*), 4, 7, and 10 with annualized management costs incurred every year (where s = 0 refers to year 1, s = 1 refers to year 4, s = 2 refers to year 7 and s = 3 refers to year

10). Note that when using equation (3) we always assume a harvest period in *yr_cur* or *yr_fut*, the next occurs $Freq_x$ years later, the next 2 $Freq_x$ years later, etc.

Alternatively, if a harvest does not take place in *yr_cur* or *yr_fut*, and instead the first one is accounted for $Freq_x$ years into the time interval T, then we use the following equation,

$$NPV_x = \sum_{s=1}^{rd\left(\frac{T_x}{Freq\_harv_x}\right)} \frac{VH_x}{\left(1+\frac{r}{100}\right)^{(Freq\_harv_x \times s)-1}} - \sum_{t=0}^{T_x-1} \left(\frac{Mait\_cost_x}{\left(1+\frac{r}{100}\right)^t}\right) \tag{5.72}$$

where "rd" means any fraction produced by T_x / $Freq\_harv_x$ is rounded down to the next integer In this case, if $Freq\_harv_x$ = 3 and T_x = 10, then x experiences a harvest period in years 3, 6, and 9 of the time interval.

The selection of T_x and $Freq_x$ require some thought. First, if timber parcel x is expected to only experience one immediate harvest period (either in the base year with equation (3) or $Freq_x$-years into the time interval with equation (6)), then set T_x = $Freq_x$ = 1. On the other hand, if parcel x is in an even-aged managed rotation, then the value of T_x can be set very high (we assume that harvests can be sustained indefinitely in such systems). However, we recommend against using large T_x values for any x for several reasons. First, in this model, timber price, harvest cost, and management cost are static over time. This may only be a reasonable assumption for short periods of time (e.g., 20 years). Second, in this model, timber management is static over time; again this may only be a reasonable assumption over short periods of time. Third, if natural forests are being transformed into plantations, a large T would require that we begin accounting for the eventual plantation harvests. This complication would make the model less tractable. Note that $Freq_x$ T_x for all x.

Finally, the net present value of timber harvest for the entire area of parcel x from the base year to T_x years later is given by TNPVx, where Parcl_areax is the area (ha^{-1}) of parcel x:

$$TNPV_x = Parcl\_area_x \times NPV_x \tag{5.73}$$

The last table entry, $BCEF_x$, is used to transform the total volume of wood removed from a parcel from *yr_cur* or *yr_fut* to T years later (TBiomassx). If $Immed\_harv_x$ = 1 then,

$$Tbiomass_x = Parcl\_area_x \times \frac{Perc\_harv_x}{100} \times Harv\_mass_x \times ru\left(\frac{T_x}{Freq\_harv_x}\right) \tag{5.74}$$

Otherwise, if $Immed\_harv_x$ = 0 then

$$Tbiomass_x = Parcl\_area_x \times \frac{Perc\_harv_x}{100} \times Harv\_mass_x \times rd\left(\frac{T_x}{Freq\_harv_x}\right) \tag{5.75}$$

and

$$TVolume_x = TBiomass_x \times \frac{1}{BCEF_x} \tag{5.76}$$

Example: Landscape with timber production in five parcels. In this example, the first two timber parcels are managed for timber production on a 45-year even-age rotation (1/45 of the stand is harvested and then replanted each year) in perpetuity, but have different mixes of species and different management costs. Each managed timber parcel is 1000 hectares. The third timber parcel has the same species mix as the second, but 1/4 of the parcel is harvested every 20 years and it will only be managed for at least another 50 years. The fourth polygon is a clear-cut of a 500 ha natural forest that is slated to become a shopping mall. The fifth parcel represents a portion of a mature, primary forest. The parcel in the larger forest that will be used for timber production is 500 ha. It will be systematically clear-cut over the next ten years and then managed as a single species plantation indefinitely (we do not account for the plantation's expected revenues in this model).

Parcel_ID	Parcl_area	Perc_harv	Freq_harv	Harv_mass	Price	Maint_cost	Harv_cost	T	Immed_harv	BCEF
1	1000	2.22	1	80	300	190	50	50	Y	1
2	1000	2.22	1	70	200	260	124	50	Y	1
3	1000	25	20	70	200	310	225	50	N	1
4	500	100	1	95	350	180	45	1	Y	1
5	500	20	2	95	400	190	105	10	Y	1

5.10.5 Running the Model

The model is available as a standalone application accessible from the Windows start menu. For Windows 7 or earlier, this can be found under *All Programs -> InVEST +VERSION+ -> Timber*. Windows 8 users can find the application by pressing the windows start key and typing "timber" to refine the list of applications. The standalone can also be found directly in the InVEST install directory under the subdirectory *invest-3_x86/invest_timber.exe*.

Viewing Output from the Model

Upon successful completion of the model, a file explorer window will open to the output workspace specified in the model run. This directory contains an *output* folder holding files generated by this model. Those files can be viewed in any GIS tool such as ArcGIS, or QGIS. These files are described below in Section *Interpreting Results*.

5.10.6 Interpreting Results

Final Results

Final results are found in the *Output* folder within the *Workspace* specified for this module.

- **Timber.shp** – The attribute table has three columns. The first column gives each timber parcel's TNPV. TNPV is the net present economic value of timber production in terms of the user-defined currency. TNPV includes the revenue that will be generated from selling all timber harvested from *yr_cur* or *yr_fut* to T years after *yr_cur* or *yr_fut* less harvest and management costs incurred during this period. Finally, all monetary values are discounted back to *yr_cur* or *yr_fut*'s present value. Negative values indicate that costs (management and harvest) are greater than income (price times harvest levels). The TBiomass column gives the total biomass (in Mg) of harvested wood removed from each timber parcel from *yr_cur* or *yr_fut* to T years after *yr_cur* or *yr_fut* (TBiomass from equation (8) or equation (9), depending on the value of Immed_harv). The TVolume column gives the total volume (m3) of harvested wood removed from each timber parcel from *yr_cur* or *yr_fut* to T years after *yr_cur* or *yr_fut* (TVolume from equation (10)).

5.10.7 References

Maass, J., P. Balvanera, A. Castillo, GC Daily, HA Mooney, P. Ehrlich, M. Quesada, A. Miranda, VJ Jaramillo, F. Garcia-Oliva, A. Martinez-Yrizar, H. Cotler, J. Lopez Blanco, A. Perez Jimenez, A. Burquez, C. Tinoco, G. Ceballos, L. Barraza, R. Ayala, and J. Sarukhan. 2005. Ecosystem services of tropical dry forests: insights from long-term ecological and social research on the Pacific Coast of Mexico. Ecology and Society 10:17.

Nunez, D., L. Nahuelhual, and C. Oyarzun. 2006. Forests and water: The value of native temperate forests in supplying water for human consumption. Ecological Economics 58:606-616.

Ricketts, TH. 2004. Tropical Forest Fragments Enhance Pollinator Activity in Nearby Coffee Crops. Conservation Biology 18:1262-1271.

Sohngen, B., and S. Brown. 2006. The influence of conversion of forest types on carbon sequestration and other ecosystem services in the South Central United States. Ecological Economics 57:698-708.

5.11 Wave Energy Production

5.11.1 Summary

Decision-makers and the public are increasingly interested in converting wave energy into electricity, with the hope that ocean waves will be a source for clean, safe, reliable, and affordable energy. The goals of the InVEST wave energy model (WEM) are to map and value the energy provisioning service provided by ocean waves and to allow for the evaluation of trade-offs that might arise when siting wave energy conversion (WEC) facilities. The WEM assesses potential wave power and harvested wave energy based on wave conditions (e.g., significant wave height and peak wave period) and technology-specific information of WEC devices (e.g., performance table and maximum capacity). The model then evaluates the net present value (NPV) of building and operating a WEC facility over its life span using economic parameters (e.g., price of electricity, discount rate, as well as installation and maintenance costs). Obtaining accurate input data and parameters for the economic valuation portion of the model is a significant challenge because there have been no commercial-scale wave energy facilities implemented to date. We recommend using the NPV values of a wave energy facility computed with the default values be used to only to make relative comparisons between sites. The outputs of the WEM provide spatially explicit information, showing potential areas for siting WEC facilities with the greatest energy production and value. This site- and device-specific information for the WEC facilities can then be used to identify and quantify potential trade-offs that may arise when siting WEC facilities. Decision-makers and stakeholders can use the WEM to better understand where to install a WEC facility with greatest harvested wave energy and least effect on coastal and ocean ecosystems and other human uses.

5.11.2 Wave Energy 3.0 Beta

Wave Energy has a 3.0 prototype which can be found in the InVEST 3.x.x folder. New features to the 3.0 version include:

- Parameters from previous runs are automatically loaded into the user interface.
- Runtime of the model has been improved.
- The projection input is no longer required and we calculate distances from the projection given by the area of interest polygon layer.
- All table inputs are now given in CSV format rather than Excel files - see sample data for correct format.

If you encounter any issues please post to the user's support forum at http://forums.naturalcapitalproject.org/

5.11.3 Introduction

Wave energy has many characteristics important to the efficient generation of electricity and is considered a potentially significant contributor to the effort to meet growing human energy demands (Barstow et al. 2008). Among various renewable energy resources, wave energy has the greatest power density and provides relatively continuous and predictable power-significant advantages for electrical grid operation (Bedard et al. 2005). The cost of electricity generated by wave energy has decreased since the 1980s and is likely to decrease further as the technology develops and the wave energy industry expands (Thorpe 1999). Considering the increasing cost of fossil fuel energy and concomitant interest in renewable energy sources, wave energy may be economically feasible in the near feature. As a consequence, decision-makers and the public are increasingly interested in converting wave energy into electricity with the hope that ocean waves will be a source for clean, safe, reliable, and affordable energy source without significant greenhouse gas emissions. With this increasing interest in wave energy as a renewable energy resource, there is a growing need for a framework to help decision-makers site wave energy facilities. The WEM we articulate here will provide planners with information that can be used to balance the harvesting of energy from waves with existing uses of marine and coastal ecosystems.

Globally, exploitable wave energy resources are approximately equal to 20% of current world electricity consumption, but their potential varies considerably by location (Cornett 2008). In addition, in areas close to the shore, "hot spots" (characterized by condensed wave energy) provide the highest potential for wave energy harvesting (Cornett and Zhang 2008; Iglesias and Carballo 2010). Therefore, indentifying wave-power-rich areas is the first step in siting a wave energy conversion (WEC) facility.

A variety of technologies for WEC devices have been proposed to capture the energy from waves, and the particular characteristics of these devices play a critical role in quantifying the amount of energy that can be captured. Therefore, the choice of WEC device is also an essential component in efficiently harvesting wave energy under different wave conditions by location.

The economic valuation of a wave energy facility can be used to compare the net benefits across sites and device-specific technologies. As with most renewable energy projects, many different factors can be included in the economic valuation. These include: the value of energy provided to the electricity grid, reduction in pollution associated with wave energy projects as compared to traditional sources, costs to those who lose access to coastal and marine locations, and environmental costs associated with the construction and operation of these facilities. In practice, including all the relevant benefits and costs, particularly those related to environmental benefits and costs, can be difficult to measure and include in a formal cost-benefit analysis. Rather than ignoring these potential impacts, we have taken a simple approach to incorporating some of this information in a framework that can be used in parallel to a formal cost-benefit analysis.

While wave energy may provide clean and renewable energy without significant greenhouse gas emissions, wave energy projects may conflict with existing ocean uses or conservation strategies for protecting marine species and habitats. WEC facilities have the potential to impact fishing opportunities, pelagic and benthic habitat, recreational activities, aesthetic views, hydrodynamic and wave environments, navigation, and the bioaccumulation of toxic materials (Boehlert et al. 2007; Nelson et al. 2008; Thorpe 1999). The severity of these potential impacts is likely to be site specific. Also, given limited experience with wave energy projects to date, there is little empirical evidence describing impacts. Therefore, identifying and evaluating the potential trade-offs associated with siting WEC facilities is an essential component of marine spatial planning and other forms of decision-making in marine and coastal environments.

The WEM presented here assesses: 1) potential wave power, 2) harvested wave energy, and 3) the net present value of a WEC facility. The outputs of the WEM provide spatially explicit information, showing potential areas for siting WEC facilities with the greatest energy production and benefits. This site- and facility-specific information then can be used to evaluate how siting a WEC facility might influence and/or change existing coastal and marine uses. For example, the WEM allows users to explore potential trade-offs by mapping and quantifying spatial competition with existing ocean uses for commercial and recreational activities (e.g., fishing, navigation, whale watching, kayaking, etc.).

5.11.4 The Model

The objective of the WEM is to help facilitate marine spatial planning in the context of wave energy projects by exploring potential costs and benefits of siting wave energy facilities. The model can run using default input data sets that are globally and regionally available or with local input data.

How it works

Potential Wave Power Resource Assessment

Wave power per unit width of wave crest length transmitted by irregular waves can be approximated as

$$P_n = \frac{\rho * g}{16} H_s^2 C_g(T_e, h) \tag{5.77}$$

where, P_n is wave power (kW/m), ρ is sea water density (1,028 kg m^{-3}), g is gravitational acceleration (9.8 m s^{-2}), H_s is significant wave height (m), and C_g is wave group velocity (m s^{-1}) as a function of wave energy period, T_e (sec), and water depth h (m) (Cornett 2008). C_g can be estimated as

$$C_g = \frac{\left(1 + \frac{2kh}{\sinh(2kh)}\right)\sqrt{\frac{g}{k}\tanh(kh)}}{2} \tag{5.78}$$

where the wave number k is calculated using a dispersion relationship expressed as a function of wave frequency ($w = 2\pi/T_e$) and water depth h:

$$w^2 = gk * \tanh(kh) \tag{5.79}$$

An iterative numerical solution scheme can be applied to solve Equation (5.79) with initial estimates of $k = w^2/(g \cdot \sqrt{tanh(w^2 \cdot h/g)})$. The wave period of measured or modeled sea states are rarely expressed as T_e, rather, they are often specified as peak wave period, T_p. Therefore, the peak energy period is estimated as $T_e = \alpha \cdot T_p$. Where, α is constant determining the shape of a wave spectrum. We use α = 0.90 as a default value assuming standard JONSWAP spectrum, which works well when sea state is dominated by waves from a single source and the spectrum is unimodal (Cornett 2008). The same assumption was also applied to global wave power resource estimation (Cornett 2008) and wave power calculations in the west coast of Canada (Cornett and Zhang 2008; Dunnett and Wallace 2009).

We prepared globally and regionally available input data layers for the potential wave power resources calculation. We used NOAA WAVEWATCH III (NWW3) model hindcast reanalysis results (version 2.22) to obtain wave characteristics defined by H_s and T_p. NWW3 spatial resolution ranges from 4 to 60 minutes depending on the global and regional grid systems. We used ETOPO1 to obtain the water depth (h), which provides 1 arc-minute global ocean bathymetry information (Amante and Eakins 2009). When using the default input data layers, model results provide the first approximation of potential wave power resources for any target area in the world. However, the spatial resolution of the model results may not be fine enough to assess wave power resources near coastal areas. So, this module will allow users to add their own wave input based on local studies (e.g., nearshore wave model results) in the next version.

Captured Wave Energy Assessment Captured wave energy can be estimated as a function of sea states and the wave energy absorption performance of a WEC device (Previsic 2004a, Previsic 2004b). A seastate is the general condition of the ocean surface and often characterized by two parameters, a significant wave height H_s and a peak period T_p. Long-term wave time-series data can be used to calculate the number of hours that each seastate occurs over a particular time period. We prepared globally and regionally available seastate tables using 3-hour interval NWW3 model results over a period of 5 years. Table 3.1 is an example of yearly occurrence of hours at each seastate bin in the west coast of Vancouver Island. In this example, a seastate with H_s = 2.5 m and T_p = 10.0 sec is most dominant, occurring 115 hours per year.

		Wave Period (T_p) in sec								
		0.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
Wave Height (H_s) in m	0.0	0	0	0	0	0	0	0	0	0
	0.1	0	0	0	0	0	0	0	0	0
	0.5	0	0	0	0	0	0	0	0	0
	1.0	0	0	0	2	0	13	12	12	6
	1.5	0	0	0	20	23	18	67	93	46
	2.0	0	0	0	12	76	34	26	131	96
	2.5	0	0	0	1	32	25	19	46	115
	3.0	0	0	0	0	6	25	16	45	72
	3.5	0	0	0	0	0	6	23	29	36
	4.0	0	0	0	0	0	1	5	15	29
	4.5	0	0	0	0	0	1	2	17	14
	5.0	0	0	0	0	0	0	1	4	5

Fig. 5.30: Occurrence of hours (hr/yr) in each seastate bin in the west coast of Vancouver Island.

The ability of a WEC device to harvest wave energy can be expressed by wave energy absorption performance that is available from WEC device manufacturers. We have conducted a literature review of WEC devices for which there is public information and prepared wave energy absorption performance tables. While these devices are technologically dated in the fast-changing offshore wave energy industry, they have undergone full-scale testing and verification in the ocean. Currently, the InVEST WEM includes as default input parameters performance tables for:

- PWP-Pelamis (Pelamis Wave Power Ltd 2010; Previsic 2004b)
- Energetech-OWC (Previsic 2004a)
- AquaBuOY (Dunnett and Wallace 2009)
- WaveDragon (Dunnett and Wallace 2009)

		Wave Period (T_e) in sec								
		0.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
Wave Height (H_s) in m	0.0	0	0	0	0	0	0	0	0	0
	0.1	0	0	0	0	0	0	0	0	0
	0.5	0	0	0	0	0	0	0	0	0
	1.0	0	0	0	11	27	50	62	64	57
	1.5	0	0	0	26	62	112	141	143	129
	2.0	0	0	0	66	109	199	219	225	205
	2.5	0	0	7	93	171	279	342	351	320
	3.0	0	0	91	180	246	402	424	417	369
	3.5	0	0	86	211	326	484	577	568	502
	4.0	0	105	216	326	394	632	616	583	585
	4.5	0	94	233	371	467	735	744	738	634
	5.0	0	259	364	469	539	750	750	750	750

Fig. 5.31: Wave energy absorption performance (kW) in each seastate bin for Pelamis.

By multiplying each cell in the annual occurrence of hours table by each corresponding cell of the wave energy absorption performance table, captured wave energy is calculated for each sea state bin. The annual captured wave energy (kWh/yr) per WEC device is calculated by summing up all the captured wave energy in each seastate bin. Some WEC devices have an ability to optimize their performance in response to site-specific conditions (Previsic 2004b) and users may need to adjust the default parameters of a WEC device or apply their own machine-dependent values for more accurate assessments of harvested wave energy.

Net Present Value Assessment We used a cost-benefit analysis framework to evaluate the construction and operation of a wave energy facility. We combined the most relevant measures of benefits (B_t) and costs (C_t) to compute the NPV for a wave energy facility located at a specific location in marine space. The NPV of a particular wave energy facility is:

$$\sum_{t=1}^{T}(B_t - C_t)(1+i)^{-t} \tag{5.80}$$

and is evaluated over a life span, T, of a WEC facility. To discount the value of future benefits and costs, a discount rate, i is required. Annual benefits are computed as the product of the price of electricity per kWh and annual captured wave energy in kWh [1]. We assume no revenue in the initial year of the project. Refer to the valuation section of the InVEST offshore wind energy model for discussion on appropriate choice of a discount rate and energy prices.

The annual costs can be broken down into initial installation costs and annual operating and maintenance costs. The initial costs of installing the wave energy devices include the following costs: 1) capital cost per installed kW, which is device dependent, 2) cost of mooring lines, 3) cost of underwater transmission cables, 4) cost of overland transmission cables [2]. Because the total costs of underwater and overland transmission cables depend on the distance of the facility to the nearest grid connection point, calculation of NPV allows users to evaluate the tradeoff between locating a facility in a particular location for its wave energy resources and the distance-dependent costs of installing the devices at that

[1] Both the discount rate and the wholesale price of electricity are user-defined inputs for which we provide example values. In many cases, fixed tariff or feed-in tariffs are being discussed to help promote development of renewable energy projects.

[2] We do not consider the costs of additional land-based infrastructure that may be required to connect an offshore facility to the grid, nor do we consider the costs of permitting a wave energy project. Costs estimates for different wave energy conversion devices were derived from Dunnett and Wallace (2009) and are given in 2006 USD$.

location. We provide default economic parameters tables for economic valuation of wave energy using three of the four machines described in the *previous section*: PWP-Pelamis, AquaBuOY and WaveDragon. All costs are expressed in 2006 US dollars and should be inflated to your study's base year of analysis.

Limitations and Simplifications

Some words of caution about limitations and simplifications of the model and guidance on model interpretation:

1. The quality of wave input data determines the accuracy of model results. So, a user needs to understand the quality of wave input data for proper interpretation of the WEM results. For example, the default wave input data are more appropriate for global and regional scale applications at 4 or 60 minutes spatial resolution. For a more detailed analysis of wave power in a region of interest, the user may want to provide wave model results obtained at a finer spatial resolution.
2. Captured wave energy indicates the yearly averaged energy absorbed per WEC device. For estimation of actual energy production from a WEC device, users may need to consider additional technology-specific information, such as device availability, power conversion efficiency, and directional factors. For some WEC devices, an increase in performance is possible without significant changes in the device structure and users may apply adjustment factors to the performance table. Please consult Previsic (2004a, 2004b) for further discussion about the estimation of actual wave energy production from a WEC facility.
3. Because there have been no commercial-scale wave energy facilities implemented to date, obtaining accurate cost data is a challenge. We provide default values for several wave energy devices that are publicly available. Because these costs may be inaccurate and/or out of date, we recommend that NPV values of a wave energy facility computed with the default values be used to only to make relative comparisons between sites. These relative comparisons will highlight that potential wave power resources and distance to the grid will have a significant influence on the estimated project cost. The magnitude of the NPV computations should be interpreted with caution.
4. The cost estimates provided are scaled for a small to moderately sized wave farm [3]. Larger farms would likely experience some cost savings from having to produce more machines, but might also require higher capacity and/or additional transmission cables. If you want to simulate the amount of energy harvested or the costs associated with a larger farm, you should carefully evaluate these factors.
5. The distance measure from a WEC facility to an underwater cable landing point is based on Euclidean metric and does not recognize any landmass within two target points. Users should be careful about distance estimation in regions with complex bathymetry.

5.11.5 Data Needs

The model uses an interface to input all required and optional model data. Here we outline the options presented to the user via the interface as well as the maps and data tables used by the model. See the appendix for detailed information on data sources and pre-processing.

Required Inputs

First we describe required inputs. The required inputs are the minimum data needed to run this model. The minimum input data allows the model to run globally without conducting economic analysis.

1. **Workspace (required).** Users are required to specify a workspace folder path. It is recommended that the user create a new folder for each run of the model. For example, by creating a folder called "runBC" within the "WaveEnergy" folder, the model will create "intermediate" and "output" folders within this "runBC" workspace.

[3] Wallace and Dunnett (2009) model 24 devices in their application.

The “intermediate” folder will compartmentalize data from intermediate processes. The model’s final outputs will be stored in the “output” folder.:

```
Name: Path to a workspace folder.  Avoid spaces.
Sample path: \InVEST\WaveEnergy\runBC
```

2. **Wave Base Data Folder (required).** Users are required to specify the path on their system to the folder with input data for the Wave Energy model. When installing InVEST, about 1GB of global Wave Watch III wave data will be included.:

```
Name: Path to a workspace folder. Avoid spaces.
Sample path (default): \InVEST\WaveEnergy\Input\WaveData\
```

3. **Analysis Area (required).** This drop down box allows users to select the scale of their analysis and instructs the model as to the appropriate wave input data. Users will also have the option of selecting an area of interest (AOI, input #7, see optional inputs below). The AOI input serves to clip these larger areas in order to perform more detailed, local analysis. If an AOI is not specified, the model will conduct wave energy calculations for the entire analysis area. There are four preset areas: West Coast of North America and Hawaii, East Coast of North America and Puerto Rico, Global (Eastern Hemisphere), and Global (Western Hemisphere):

```
File type: drop down options
Sample (default): West Coast of North America and Hawaii
```

4. **Machine Performance Table (required).** This table indicates a machine’s “performance”, or its ability to capture wave energy given seastate conditions. The first row indicates wave period bins (Tp) in seconds while the first column indicates wave height bins (Hs) in meters. The remaining numbers in the table indicates captured wave energy for the given seastate condition defined by wave height (Hs) and period (Tp).:

```
Table Names: File can be named anything, but no spaces in the name
    File type: *.csv
    Sample data set:  \InVEST\WaveEnergy\Input\Machine_Pelamis_Performance
```

		Wave Period (T_p) in sec								
		0.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
Wave Height (H_s) in m	0.0	0	0	0	0	0	0	0	0	0
	0.1	0	0	0	0	0	0	0	0	0
	0.5	0	0	0	0	0	0	0	0	0
	1.0	0	0	0	11	27	50	62	64	57
	1.5	0	0	0	26	62	112	141	143	129
	2.0	0	0	0	66	109	199	219	225	205
	2.5	0	0	7	93	171	279	342	351	320
	3.0	0	0	91	180	246	402	424	417	369
	3.5	0	0	86	211	326	484	577	568	502
	4.0	0	105	216	326	394	632	616	583	585
	4.5	0	94	233	371	467	735	744	738	634
	5.0	0	259	364	469	539	750	750	750	750

5. **Machine Parameters Table (required).** This table indicates a machine’s maximum capacity and limits (wave height and period) to capturing wave energy given seastate conditions.:

```
Table Names: File can be named anything, but no spaces in the name
    File type: *.csv
    Sample data set: \InVEST\WaveEnergy\Input\Machine_Pelamis_Parameter
```

6. **Global Digital Elevation Model (DEM) (required).** A bathymetric raster layer is required to calculate ocean depths in meters. This information is incorporated into potential wave power calculation and the economic analysisvaluation to determine the cost to send mooring cables to the ocean floor before running them to landing

NAME	VALUE	NOTE
CapMax	750	Maximum capacity of device [kW].
HsMax	10.0	Upper limit of wave height for device operation [m]: this device shuts down when wave height is bigger than HsMax.
TpMax	20.0	Upper limit of wave period for device operation [sec]: this device shuts down when wave period is longer than TpMax.

points. If the user specifies a raster input that doesn't cover the entire AOI, then wave output results outside this coverage will not include wave power calculations. To ensure the model runs properly, make sure this input covers the analysis area specified in input #2 and #7. The default bathymetry data, global_dem, provides 1 arc-minute global bathymetry data. If you are using wave input data coarser than 1arc1 arc-minute resolution, we recommend using the global demDEM data.:

```
Name: File can be named anything, but no spaces in the name and less than 13 characters
Format: GIS raster file (e.g., ESRI GRID or IMG) with depth information in meters
Sample data set (default): \InVEST\Base_Data\Marine\DEMs\global_dem
```

Optional Inputs

The next series of inputs are optional, but may be required depending on other decision inputs.

7. **Results Suffix** Appends a suffix to outputs to track model runs.
8. **Area of Interest (AOI) (required for economic valuation).** If you would like to further narrow your analysis area, you can create a polygon feature layer that defines your area of interest. It instructs the model where to clip the input data and defines the exact extent of analysis. This input is only required, however, when running the economic valuation. At the start, the model will check that the AOI is projected in meters and the datum is WGS84. If not, it will stop and provide feedback.:

```
Name: File can be named anything, but no spaces in the name
File type: polygon shapefile (.shp)
Sample path: \InVEST\WaveEnergy\Input\AOI_WCVI.shp
```

9. **Compute Economic Valuation?** By checking this box, users will instruct the model to run the economic valuation of the model. Currently, valuation is only permitted for runs where there is an AOI (input #8). Additionally, the following inputs (#10-12) must be also be specified in order to output economic analysis.
10. **Grid Connection Points File (optional, but required for economic valuation).** When running the economic analysis, you must provide a .csv table that specifies locations where machine cables would reach land and eventually the energy grid. A point ID, latitude and longitude coordinates and the type of point are required. The model will use this input to create a point feature class and project it based on the projection of the AOI input #4.:

```
Table Names: File can be named anything, but no spaces in the name
File type: *.csv
Sample data set: \InVEST\WaveEnergy\Input\LandGridPts_WCVI.csv
```

When filling out the tables with your own data, make sure to:

- Specify latitude and longitude in decimal degrees (as shown below)
- Only include the words "LAND" or "GRID" in the "TYPE" column. Use the "TYPE" field to differentiate between the two landing types.

ID	LAT	LONG	TYPE	LAT	LONG	TYPE	LOCATION
1	48.92100	-125.54200	LAND	48.99700	-125.58300	GRID	Ucluelet
2	49.13900	-125.91500	LAND	49.09800	-125.85600	GRID	Tofino

11. **Machine Economic Table (optional, but required for economic valuation).** When running the economic analysis, the user must enter a table that includes the price of electricity, machine setup and cable costs, and other valuation parameters for net present value (NPV) calculations. Sample data for three different machines are available in InVEST. Sample costs are given in 2006 USD$:

```
Table Names: File can be named anything, but no spaces in the name
File type: *.csv
Sample data set: \InVEST\WaveEnergy\Input\Machine_Pelamis_Economic.csv
```

NAME	VALUE	NOTE
CapMax	750	Maximum capacity of device [kW].
Cc	3671	Capital cost per installed [$/kW].
Cml	20	Cost of mooring lines [$ per m].
Cul	101609	Cost of underwater transmission line [$ per km].
Col	64499	Cost of overland transmission line [$ per km].
Omc	0.042	Operating & maintenance cost [$ per kWh].
P	0.200	Price of electricity [$ per kWh].
R	0.080	Discount rate
Smlpm	3.0	Slack-moored (i.e. requires length of 3 * depth); 3 slack lines per machine required.

12. **Number of Machine Units (optional, but required for economic valuation).** When running the economic analysis, the user must enter an integer value for the number of devices per wave energy facility. This value is used for determining total energy generated during the life span (25 years) of a wave energy conversion facility.

    To determine a reasonable number of machines to enter, we recommend that the user divide the maximum capacity of the machine (see input #5) by the desired amount of energy captured. For example, if the user desires 21,000 kW of captured wave energy, then the wave energy farm would have 28 Pelamis (maximum capacity is 750kW), or 84 AquaBuoy (maximum capacity is 250kW), or 3 WaveDragon (maximum capacity is 7000kW).

5.11.6 Running the Model

The model is available as a standalone application accessible from the Windows start menu. For Windows 7 or earlier, this can be found under *All Programs -> InVEST +VERSION+ -> Wave Energy*. Windows 8 users can find the application by pressing the windows start key and typing "wave" to refine the list of applications. The standalone can also be found directly in the InVEST install directory under the subdirectory *invest-3_x86/invest_wave_energy.exe*.

Viewing Output from the Model

Upon successful completion of the model, a file explorer window will open to the output workspace specified in the model run. This directory contains an *output* folder holding files generated by this model. Those files can be viewed in any GIS tool such as ArcGIS, or QGIS. These files are described below in Section *Interpreting Results*.

5.11.7 Interpreting Results

Model Outputs

The following is a short description of each of the outputs from the Wave Energy model. Each of these output files is automatically saved in the "Output" & "Intermediate" folders that are saved within the user-specified workspace directory:

Output Folder

- Output\wp_kw & Output\wp_rc
  - These raster layers depict potential wave power in kW/m for the user-specified extent. The latter ("_rc") is the former reclassified by quantiles (1 = < 25%, 2 = 25-50%, 3 = 50-75%, 4 = 75-90%, 5 = > 90%). The ("_rc") raster is also accompanied by a csv file that shows the value ranges for each quantile group as well as the number of pixels for each group.
  - The potential wave power map indicates wave power resources based on wave conditions. These often provide the first cut in the siting process for a wave energy project.
- Output\capwe_mwh & Output\capwe_rc
  - These raster layer depict captured wave energy in MWh/yr per WEC device for the user-specified extent. The latter ("_rc") is the former reclassified by quantiles (1 = < 25%, 2 = 25-50%, 3 = 50-75%, 4 = 75-90%, 5 = > 90%). The ("_rc") raster is also accompanied by a csv file that shows the value ranges for each quantile group as well as the number of pixels for each group.
  - The captured wave energy map provides useful information to compare the performance of different WEC devices as a function of site-specific wave conditions.
- Output\npv_usd & Output\npv_rc
  - These raster layers depict net present value in thousands of $ over the 25 year life-span of a WEC facility for the user-specified extent. The latter ("_rc") is positive values of the former reclassified by quantiles (1 = < 25%, 2 = 25-50%, 3 = 50-75%, 4 = 75-90%, 5 = > 90%). The ("_rc") raster is also accompanied by a csv file that shows the value ranges for each quantile group as well as the number of pixels for each group.
  - The NPV map indicates the economic value of a WEC facility composed of multiple devices. A positive value indicates net benefit; a negative value indicates a net loss. Such information can be used to locate potential areas where a wave energy facility may be economically feasible.
  - These are only an output if you have chosen to run economic valuation.
- Output\LandPts_prj.shp and GridPt_prj.shp
  - These feature layers contain information on underwater cable landing location and power grid connection points.
  - The landing and grid connection points provide useful information for interpreting the NPV map.
  - It is only an output if the user chooses to run the economic valuation.
- Parameters_[yr-mon-day-min-sec].txt
  - Each time the model is run a text file will appear in the workspace folder. The file will list the parameter values for that run and be named according to the date and time.
  - Parameter log information can be used to identify detailed configurations of each of scenario simulation.

Intermediate Folder

- intermediate\WEM_InputOutput_Pts.shp
  - These point layers from the selected wave data grid are based on inputs #2-4.
  - They contain a variety of input and output information, including:
    - I and J – index values for the wave input grid points
    - LONG and LAT – longitude and latitude of the grid points
    - HSAVG_M – wave height average [m]
    - TPAVG_S – wave period average [second]
    - DEPTH_M – depth [m]
    - WE_KWM – potential wave power [kW/m]
    - CAPWE_MWHY – captured wave energy [MWh/yr/WEC device]
    - W2L_MDIST – Euclidean distance to the nearest landing connection point [m]
    - LAND_ID – ID of the closest landing connection point that is closest
    - L2G_MDIST – Euclidean distance from LAND_ID to the nearest power grid connection point [m]
    - UNITS – number of WEC devices assumed to be at this WEC facility site
    - CAPWE_ALL – total captured wave energy for all machines at site [MWh/yr/WEC facility]
    - NPV_25Y – net present value of 25 year period [thousands of $]
  - The model outputs in raster format are interpolated results based on these point data. So, you can use this point information to explore the exact values of essential inputs and outputs at wave input data point locations.
- intermediate\GridPt.txt and LandPts.txt + These text files log records of the grid and landing point coordinates. + This is only an intermediate output if you choose to run economic valuation.

5.11.8 Case Example Illustrating Results

The following example illustrates the application of the wave energy model to the west coast of Vancouver Island (WCVI). The figures and maps are for example only, and are not necessarily an accurate depiction of WCVI. In this example, we use input data layers including:

1. Wave base data = West Coast of North America with 4-minute resolution 2. Area of Interest = AOI_WCVI.shp 3. WEC device = Pelamis 4. Digital Elevation Model = global_dem 5. Landing and Power Grid Connection Points = LandGridPts_WCVI.shp 6. Number of Machine Units = 28 7. Projection = WGS 1984 UTM Zone 10N.prj

In order to generate a grid-scale power producing facility, it is necessary to capture a minimum of 10 kW/m of wave power (Spaulding and Grilli 2010). Along the WCVI, this threshold is generally met, with the annual mean wave power >10 kW/m in most areas. Wave power gradually increases offshore. Approximately 20 kW/m wave power is available within 10 km of the shore, but the maximum wave power, 30-40 kW/m, is available 20-60 km offshore where depth is > 150 m.

Captured wave energy in this example is calculated based on Pelamis devices with 750 kW power rating. The overall patterns of the captured wave energy are similar to those of potential wave power. A Pelamis device located at the 50-70 m depth contour produces approximately 2,000-2,300 MWh/yr of energy. Assuming 15 MWh/yr energy use per household in the WCVI (Germain 2003), each Pelamis unit produces enough energy to support 133-153 households.

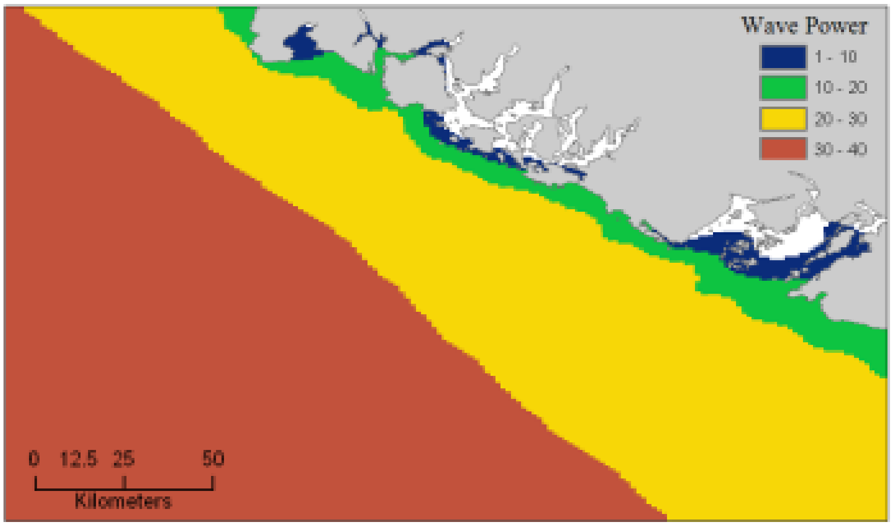


Fig. 5.32: Wave power potential (kW/m) in the west coast of Vancouver Island.

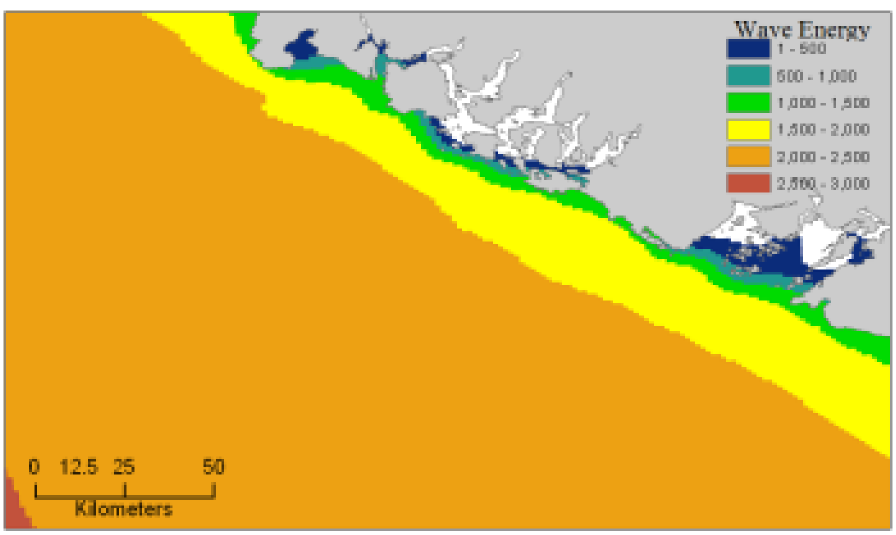


Fig. 5.33: Captured wave energy (MWh/yr) using a Pelamis device with a 750 kW power rating.

For the economic valuation of harvested wave energy, we calculate and map NPV over the 25-yr life-span of a WEC facility. For this example model run, each of the WEC facilities is composed of 28 Pelamis devices. We used an estimate of $100,000 for the underwater cable cost and 20 cents/kW for the price of electricity. Positive NPV occurs from 5-10 km offshore from the shoreline. It increases offshore and the highest NPV (the top 20% of all calculated NPV values ($4668k - $7307k)) occurs between 25-90 km from the shore.

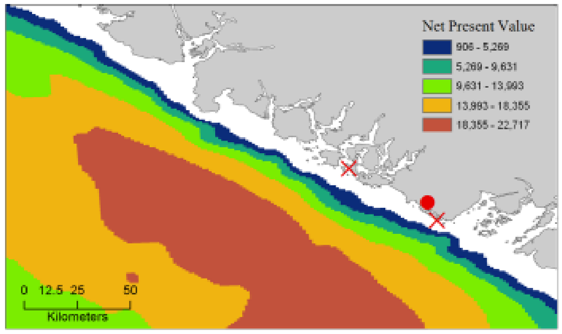


Fig. 5.34: Net present value (thousand $) over a 25-yr life-span, using $100,000 per km for the cost of underwater transmission cables. Two underwater cable landing points are located in Tofino and Ucluelet (×) and the power grid connection point is located in Ucluelet (o). Each of the WEC facilities is composed of 28 Pelamis devices and the price of electricity is set at 20 cents per kW.

Because there have been no commercial-scale wave energy facilities implemented to date, large uncertainties exist in the economic parameters. In particular, the cost of underwater transmission cables is highly uncertain, ranging from $100,000 to $1,000,000 per km. The NPV uses a lower bound of $100,000 per km for the cable cost. When we use a median cost of underwater transmission cables ($500,000 per km), the area with a positive NPV is significantly reduced.

In this example, positive NPV only occurs within a 50 km radius around the two underwater cable landing points in Tofino and Ucluelet. The upper 20% NPV exists between 10-40 km distances from the two landing points. When the upper bound ($1,000,000 per km) of transmission cable costs is used, no positive NPV exist in the WCVI. Considering uncertainties in economic parameters, users should be cautious in interpreting the magnitude of the NPV. We recommend that the NPV of a wave energy facility computed with the default values be used only to make relative comparisons between sites.

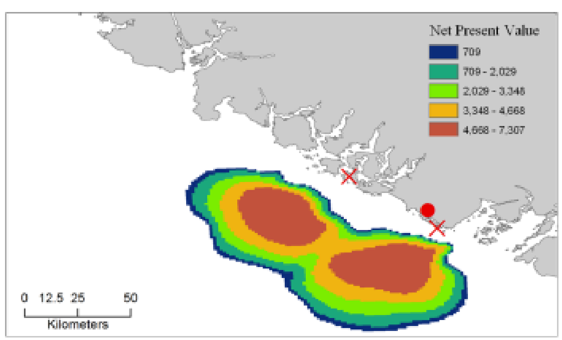


Fig. 5.35: Net present value (thousand $) over a 25-yr life-span, using $500,000 per km for the cost of underwater transmission cables. Two underwater cable landing points are located in Tofino and Ucluelet (×) and power grid connection point is located in Ucluelet (o). Each of the WEC facilities is composed of 28 Pelamis devices. The price of electricity is set at 20 cents per kW.

5.11.9 Appendix A

Data Sources

This is a rough compilation of data sources and suggestions for finding, compiling, and formatting data. This section should be used for ideas and suggestions only. We will continue to update this section as we learn about new data sources and methods.

- Wave data: significant wave height (H_s) and peak wave period (T_p)
  - Global ocean wave buoy data are available from NOAA's National Data Buoy Center (http://www.ndbc.noaa.gov/). Although ocean wave buoy provides the most accurate wave time series data, their spatial resolution is very coarse and it may not be appropriate for local scale analysis.
  - NOAA's National Weather Service provides WAVEWATCH III model hindcast reanalysis results (http://polar.ncep.noaa.gov/waves/index2.shtml). The spatial resolution of the model results ranges from 4 to 60 minutes depending on the global and regional grid systems. The model outputs have been saved at 3-hour interval from 1999 to the present. The model results have been validated with ocean buoy data at many locations and provide good quality wave information.
- Water depth
  - NOAA's National Geophysical Data Center (NGDC) provides global bathymetry data with various spatial resolutions at http://www.ngdc.noaa.gov/mgg/bathymetry/relief.html.
  - ETOPO1 is a 1 arc-minute global relief model of Earth's surface that integrates land topography and ocean bathymetry. It was built from numerous global and regional data sets, and is available in "Ice Surface" (top of Antarctic and Greenland ice sheets) and "Bedrock" (base of the ice sheets) versions. NGDC also provides regional and other global bathymetry datasets.
- Wave energy absorption performance
  - EPRI wave energy conversion project provides a review of several WEC devices: http://oceanenergy.epri.com/waveenergy.html
  - Recent updates on technology may be available from the WEC device manufactures.
    - PWP-Pelamis: http://www.pelamiswave.com/
    - AquaBuOY: http://www.finavera.com/
    - WaveDragon: http://www.wavedragon.net/
    - DEXAWAVE: http://www.dexawave.com/

5.11.10 References

Amante, C., and B. W. Eakins. 2009. ETOPO1 1 Arc-minute global relief model: procedures, data sources and analysis, p. 19. NOAA Technical Memorandum NESDIS NGDC-24.

Barstow, S., G. Mørk, D. Mollison, and J. Cruz. 2008. The wave energy resource, p. 94-131. In J. Cruz [ed.], Ocean Wave Energy: current status and future prepectives. Springer.

Bedard, R., G. Hagerman, M. Previsic, O. Siddiqui, R. Thresher, and B. Ram. 2005. Offshore wave power feasibility demonstration project: final summary report, p. 34. Electric Power Research Institute Inc.

Boehlert, G. W., G. R. Mcmurray, and C. E. Tortorici. 2007. Ecological effects of wave energy development in the Pacific Nothwest, p. 174. U.S. Dept. Commerce, NOAA Tech. Memo.

Cornett, A., and J. Zhang. 2008. Nearshore wave energy resources, Western Vancouver Island, B.C., p. 68. Canadian Hydraulics Centre.

Cornett, A. M. 2008. A global wave energy resource assessment. Proc. ISOPE 2008.

Dunnett, D., and J. S. Wallace. 2009. Electricity generation from wave power in Canada. Renewable Energy 34: 179-195.

Germain, L. A. S. 2003. A case study of wave power integration into the Ucluelet area electrical grid. Master Thesis. University of Victoria.

Iglesias, G., and R. Carballo. 2010. Wave energy and nearshore hot spots: the case of the SE Bay of Biscay. Renewable Energy 35: 2490-2500.

Nelson, P. A. and others 2008. Developing wave energy in coastal California: potential socio-economic and environmental effects, p. 182. California Energy Commission, PIER Energy-Related Environmental Research Program, and California Ocean Protection Council.

Pelamis Wave Power Ltd. 2010. Pelamis Wave Power. http://www.pelamiswave.com/.

Previsic, M. 2004a. System level design, performance and costs - San Francisco California Energetech offshore wave power plant. EPRI.

—. 2004b. System level design, performance and costs for San Francisco California Pelamis offshore wave power plant, p. 73. EPRI.

Spaulding, M. L., and A. Grilli. 2010. Application of technology development index and principal component analysis and cluster methods to ocean renewable energy facility siting. Marine Technology Society Journal 44: 8-23.

Thorpe, T. W. 1999. A brief review of wave energy, p. 186. The UK department of trade and industry.

Wilson, J. H., and A. Beyene. 2007. California wave energy resource evaluation. Journal of coastal research 23: 679-690.

5.12 Offshore Wind Energy Production

5.12.1 Summary

Offshore wind energy is gaining interest worldwide, with 5,400 megawatts (MW) installed as of January 2013 and a growth rate around 25% per year (GWEC, 2013). Consistently higher offshore winds and proximity to coastal load centers serve as two of the major reasons wind energy developers are looking offshore. The goal of the InVEST offshore wind energy model is to provide spatial maps of energy resource availability, energy generation potential, and (optionally) energy generation value to allow users to evaluate siting decisions, use tradeoffs, and an array of other marine spatial planning questions. The model was developed to allow maximum flexibility for the user, in that it can be run with default data and parameters, but it can just as easily be updated with new turbine and foundation information, grid connection information, and parameter values that fit the user's context. Model outputs include wind power potential, energy generation, offset carbon emissions, net present value, and levelized cost of energy, all given at the farm level. Peer-reviewed references for this model are http://dx.doi.org/10.1016/j.aquaculture.2014.10.035 for the financial portion of the model and http://dx.doi.org/10.1016/j.marpol.2015.09.024 for the physical portion.

5.12.2 Introduction

This wind energy model provides an easily replicable interface to assess the viability of wind energy in your region under different farm design scenarios. The outputs are raster maps, whose point values represent the aggregate value of a farm centered at that point. This allows for detailed analysis of siting choices at a fine scale, though it comes at the cost of assuming that conditions are sufficiently symmetric around the center point so that the center point represents the median conditions of all turbines in the farm. Since the user can select the number of turbines for the farm, and the raster maps do not give an indication of farm size, the model also outputs a representative polyline polygon at a randomly selected wind data point that indicates the size of the farm.

To run the model, you are asked to supply information into the graphical user interface. This includes information about the type of turbine, number of turbines, the area of interest, etc. To minimize the set of required inputs, the model includes default data in *.csv* tables on two common offshore wind turbines: 3.6 MW and 5.0 MW. Also included is a table of less commonly changed default values used to parameterize various parts of the model, called the "Global Wind Energy Parameters" file. These *.csv* files are required inputs, and may be modified if alternate values are desired by directly editing the files using a text editor or Microsoft Excel. When modifying these files, it is recommended that the user make a copy of the default *.csv* file so as not to lose the original default values.

5.12.3 The Model

Wind Energy Potential

The wind energy model estimates wind power density (wind power potential) to identify offshore areas with high energy potential. The wind power density $PD(Wm^{-2})$ at a certain location can be approximated as a function of wind statistics (Elliott et al., 1986)

$$\frac{1}{2}\rho\sum_{j=1}^{c} f(V_j)V_j^3 \tag{5.81}$$

where, ρ is mean air density ($kg\,m^{-3}$), j is the index of wind speed class, c is the number of wind speed classes, V_j is wind speed of the jth class (ms^{-1}), and $f(V_j)$ is probability density function of V_j. Two probability distributions are commonly used in wind data analysis: 1) the Rayleigh and 2) the Weibull distributions (Manwell et al. 2009). The Weibull distribution can better represent a wider variety of wind regimes (Celik 2003; Manwell et al. 2009), and is given as

$$f(V_j) = \frac{k}{\lambda}\left(\frac{V_j}{\lambda}\right)^{k-1} e^{-\left(\frac{V_j}{\lambda}\right)^k} \tag{5.82}$$

where, k and λ are the shape and scale factors, respectively. The shape factor, k, determines the shape of the Weibull probability density function (Figure 5.36). The probability density function shows a sharper peak as k increases,indicating that there are consistent wind speeds around the mean wind speed. On the other hand, the function becomes smoother as k decreases, indicating more variation in wind speed and more frequent low and high wind speeds. We used a MATLAB function, *wblfit*, to estimate k and λ, which returns the maximum likelihood estimates of the parameters of the Weibull distribution given the values in the wind time series data. For more details of *wblfit* function, please consult http://www.mathworks.co.kr/kr/help/stats/wblfit.html.

Wind power density is calculated at the hub height Z (m) of a wind turbine (Figure 5.36), which means all variables in (5.81) and (5.82) need to be converted into the appropriate value at hub height. Mean air density ρ was estimated as $\rho = 1.225 - (1.194 \cdot 10^{-4})Z$, which approximates the U.S. Standard Atmosphere profile for air density (National Oceanic and Atmospheric Administration, 1976). We applied the wind profile power law to estimate wind speed (V) at hub height Z (Elliott et al., 1986).

$$\frac{V}{V_r} = \left(\frac{Z}{Z_r}\right)^{\alpha} \tag{5.83}$$

where V is wind speed (ms^{-1}) at the hub height Z (m) of a wind turbine, and V_{ris} wind speed (ms^{-1}) at the reference height Z_r (m) where wind data are obtained. α is power law exponent, which is an empirically derived coefficient and varies with the stability of the atmosphere. For neutral stability condition, α is approximately 1/7 (0.143) for land surfaces, which is widely applicable to adjust wind speed on land (Elliott et al., 1986). The power law exponent has different value on ocean surfaces. Hsu et al (1994) found that $\alpha = 0.11 \pm 0.03$ for ocean surface under near-neutral atmospheric stability conditions. The wind energy model uses $\alpha = 0.11$ as a default value to adjust wind speed on the ocean surface. The wind profile of the atmospheric boundary layer can be approximated more accurately using the log wind profile equation that accounts for surface roughness and atmospheric stability (Manwell et al. 2009).

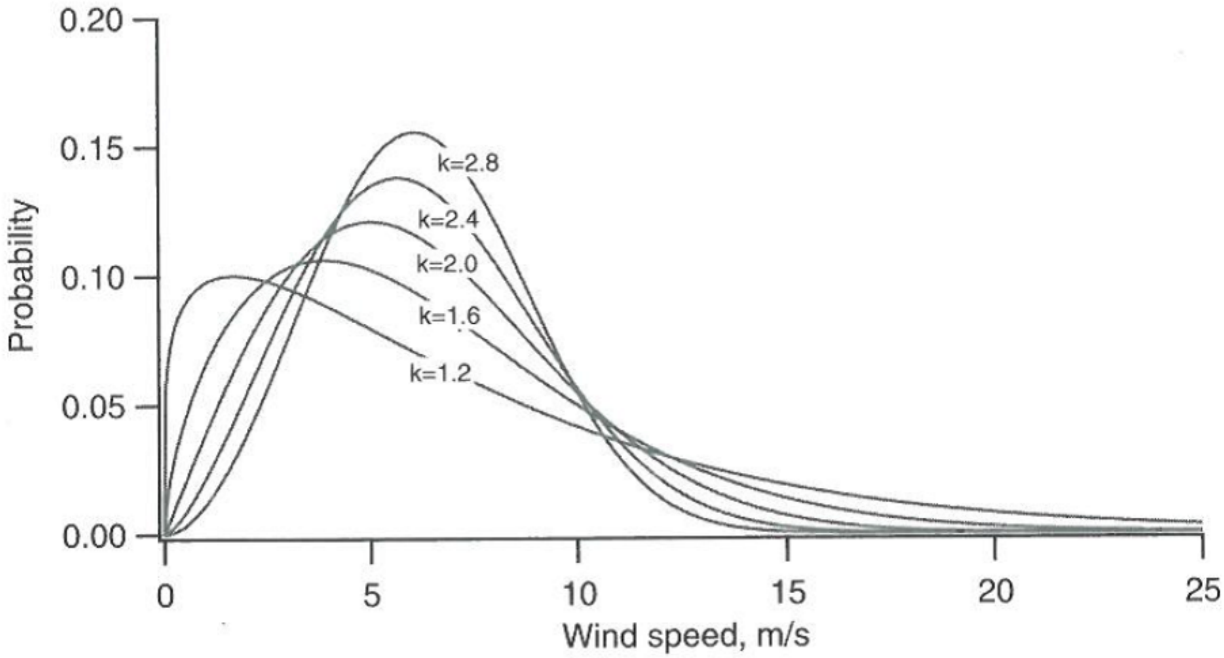


Fig. 5.36: Example of Weibull probability density function with various shape factors (k), where mean wind velocity = $6ms^{-1}$ (Manwell et al., 2009).

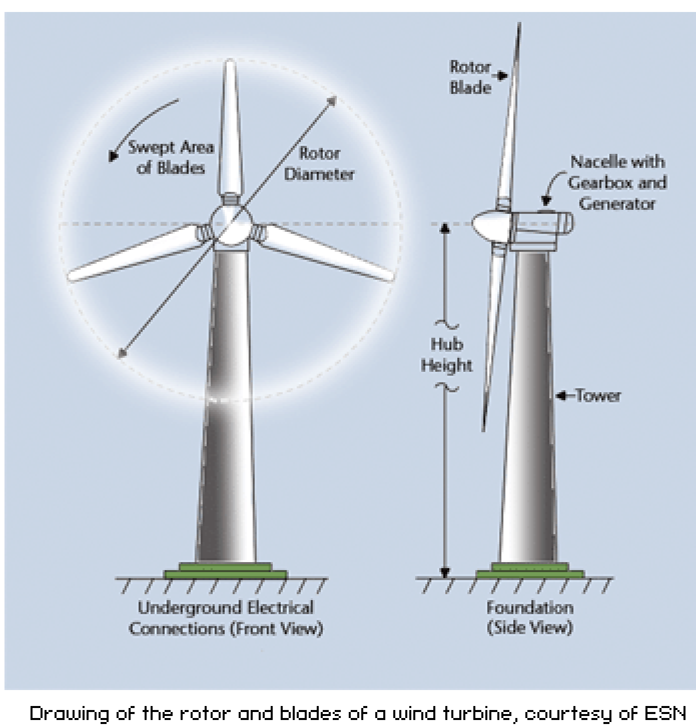


Fig. 5.37: A schematic diagram of a wind turbine (http://www.daviddarling.info/encyclopedia/H/AE_hub_height.html)

Wind power density (PD) outputs provide suitability information for a wind energy development project in terms of wind resource. Pacific Northwest Laboratories categorized wind power density and wind speed into seven classes based on United States wind atlas (Figure 5.38) (Manwell et al. 2009). Areas designated as class 4 or greater are considered to be suitable for most wind energy development. Class 3 areas are suitable for wind energy development if large turbines are used. Class 1 and 2 are rarely considered as suitable areas for wind energy development in terms of energy potential. Wind resources vary considerably over space and a more detailed categorization of wind power density for five topographical conditions was developed in Europe, which includes sheltered terrain, open plain, sea coast, open sea, hills and ridges (Figure 5.38) (Manwell et al. 2009). The wind resource classification for sea coast and open sea may provide better information on the suitability of offshore wind energy projects.

Map Color	Sea Coast (50 m height)		Open Sea (50 m height)	
	Wind Power Density (W/m^2)	Wind Speed (ms^{-1})	Wind Power Density (W/m^2)	Wind Speed (ms^{-1})
Blue	0-150	0.0-5.0	0-200	0.0-5.5
Green	150-250	5.0-6.0	200-400	5.5-7.0
Orange	250-400	6.0-7.0	400-600	7.0-8.0
Red	400-700	7.0-8.5	600-800	8.0-9.0
Purple	> 700	> 8.5	> 800	> 9.0

Fig. 5.38: Wind power density (PD) and wind speed classes based on European wind atlas (Modified from Table 2.6 in Manwell et al. 2009).

Energy Generation

The amount of energy harvestable from a wind turbine in a particular location depends on the characteristics of the wind turbine as well as wind conditions (Pallabazzer 2003; Jafarian & Ranjbar 2010). The wind energy model quantifies the harvestable energy based on the output power curve of a wind turbine and wind speed statistics. Figure 5.39 shows an output power curve of a wind turbine (pitch control type). The wind turbine starts to generate power at the cut-in wind speed (V_cin). The output power increases up to the rated power (Prate) as wind speed increases to the rated wind speed (V_rate). The wind turbine keeps producing the maximum power (i.e., Prate) until wind speed reaches the cut-out wind speed (V_cout). If wind speed increases beyond the cut-out wind speed, the wind turbine stops generating power for safety purposes. Currently, more than 74 offshore wind farms are operating globally and technology specific information of the wind turbine at each wind farm are available at LORC Knowledge (2012).

To provide flexibility for a variety of different turbine types without requiring the user to manually enter in a power curve, we estimate the output power P (kW) of a wind turbine using a polynomial modeling approach (Jafarian & Ranjbar 2010):

$$P(V) = \begin{cases} 0 & V < V_{cin} \text{ or } V > V_{cout} \\ P_{rate} & V_{rate} < V < V_{cout} \\ (V^m - V_{in}^m)/(V_{rate}^m - V_{in}^m) & V_{cin} \leq V \leq V_{rate} \end{cases} \tag{5.84}$$

where, m is an exponent of the output power curve (usually 1 or 2). Using this approach, the energy output, O (MWh), generated by a wind turbine can be calculated using

$$O = nday \cdot \frac{\rho}{\rho_0} P_{rate} \left(\int_{V_{cin}}^{V_rate} \frac{V^m - V_{cin}^m}{V_r^m - V_{cin}^m} f(V) dV + \int_{V_{rate}}^{V_{cout}} f(V) dV \right) (1 - lossrate) \tag{5.85}$$

where, $nday$ is the number of days for energy output (e.g. $nday = 365$ days for annual energy output), ρ_0 is air density of standard atmosphere (e.g. $1.225 kgm^{-3}$ for U.S. standard atmosphere air density at sea level), and $lossrate$

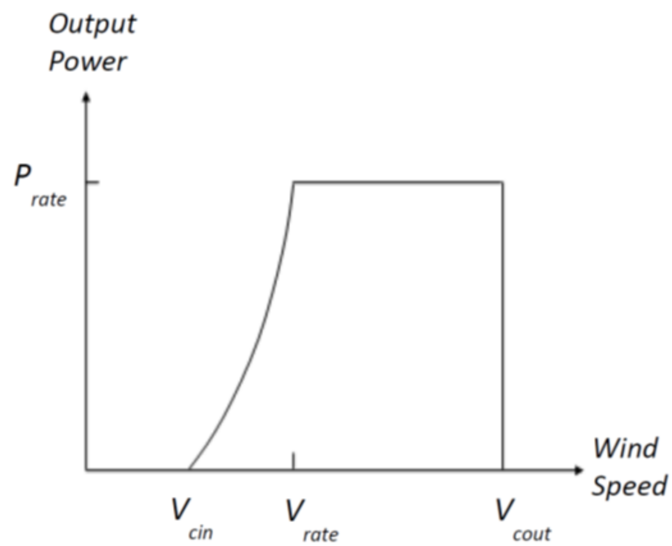


Fig. 5.39: Output power (P) curve of a wind turbine (pitch control type) as a function of wind speed (V) (Modified from Fig.1 in Pallabazzer 2003)

is a decimal value which represents energy losses due to a combination of downtime, power conversion efficiency, and electrical grid losses (default value is .05). All of these parameters are included in the global parameters *.csv* file and may be changed by the user from their defaults. Total farm energy output is equal to the individual turbine output multiplied by the number of turbines, n,

$$E = nO \tag{5.86}$$

The InVEST software comes with default technical and financial information about two common turbine sizes, the 3.6 MW and 5.0 MW turbines. The information for each turbine is given in *.csv* files in the *Input* directory and is a required input into the model. The user can use the default data, edit a file, or create a new file to assess different turbine sizes or update specific characteristics. The files must retain the same format - only parameter values may safely be modified. It is recommended to save edits as new *.csv* files rather than overwriting the default data.

Offset Carbon

Since wind turbines create no greenhouse gasses when generating energy, the user may be interested in assessing the amount of carbon dioxide emissions avoided by building a wind farm versus a conventional energy generation plant. To translate carbon-free wind power to a representative amount of annual avoided $\mathrm{CO_2}$ emissions, we use the following default conversion factor: $6.8956 \cdot 10^{-4} \mathrm{metric\ tons\ CO_2}/kWh$

This is obtained from the EPA (http://www.epa.gov/cleanenergy/energy-resources/refs.html) and is based on 2007 data. See their website for limitations of this approach. The parameter is representative of the carbon emitted by the energy portfolio of the United States and may not be appropriate for your context. This value is changeable in the global parameters *.csv* file.

Value of Power

The value of wind power is measured as the discounted pre-tax net revenue from power generation that would accrue to a wind power developer/operator over the expected lifetime of a wind farm. The Net Present Value (http://en.wikipedia.org/wiki/Net_present_value) (NPV) of energy for a given wind farm is:

$$NPV = \sum_{t=1}^{T} (R_t - C_t)(1+i)^{-t} \tag{5.87}$$

Where R_t is the gross revenue collected in year t, and C_t are the aggregate costs in year t. math:T represents the expected lifetime of the facility, and i represents the discount rate (http://en.wikipedia.org/wiki/Discount_rate) or weighted average cost of capital (WACC, http://en.wikipedia.org/wiki/Weighted_average_cost_of_capital). Both T and i can be changed by the user; T can be found in the global parameters *.csv* file and i is entered in the valuation section of the user interface. For projects that are financed by both debt and equity and where there is a significant amount of risk associated with establishing and maintaining the projected stream of revenues, WACC is a more appropriate method for establishing the time value of money. As this parameter enters into the calculation in the same way as a discount rate would, if you prefer you can input an appropriate discount rate and interpret the results accordingly. We do not supply a default value, but Levitt et al. (2011) suggest a WACC value of .116 based on a comprehensive analysis of industry specific discount rates and different debt/equity structures in Europe and the U.S. This is higher than discount rates typically used elsewhere, such as in standard cost benefit analysis, so you may find your application justifies a different rate.

Annual gross revenue is calculated by multiplying the price per kWh, s, by the annual amount of kWh supplied to the grid by a wind farm, E_t, thus $R_t = sE_t$. It is assumed that energy is not collected in the first year during the construction phase.

Costs can be separated into one-time capital costs and ongoing operations and management costs. During the construction phase, expenditures are made on turbines, foundations, electrical transmission equipment, and other miscellaneous costs associated with development, procurement, and engineering. At the end of the farms usable lifetime, the firm must remove their equipment. The default information supplied is based on an extensive review of peer-reviewed publications, industry reports, and press releases. This information is summarized below.

Turbines

Turbines and foundations are modeled with unit costs. We have supplied cost data on 3.6 MW and 5.0 MW class turbines as well as monopile and jacketed foundations, though you may enter your own turbine- or foundation-specific information. Note all default costs below are given in 2012 US dollars. Assuming one foundation per turbine, the total cost of turbines and foundations is simply the number of wind turbines multiplied by the unit cost. Table 1 gives a summary of existing turbine costs.

Farm	Location	Capacity	# of Turbines	Total MW	Unit Cost ($mil)
Riffgat	UK	3.6	30	108	6.3
Sheringham Shoal	UK	3.6	88	317	5.65
Greater Gabbard	UK	3.6	140	504	6.03
Butendiek	Germany	3.6	80	288	5.54
London Array	UK	3.6	175	630	6.29
Amrumbank	Germany	3.6	80	288	6.41
Global Tech 1	Germany	5	80	400	10.4
Borkum 2	Germany	5	40	200	10.6

Table 1: Turbine costs.

Foundations

This model can flexibly include valuation for both foundation-based and floating turbine designs. This is accomplished by letting the user enter the appropriate unit cost information for their farm design. Outputs are constrained by user-editable depth and distance parameters, so it is important to adjust these to reflect the appropriate technological constraints of your design choice. Foundation-based turbines have conventionally been limited to a depth of around 60 meters.

Foundation cost information is relatively difficult to come by. Monopile foundations are the most common foundation type and are typically mated to 3.6 MW turbines. Ramboll, a major foundation manufacturer, estimates that monopile foundations with a 3.6 MW turbine are $2 million per foundation. Monopile costs at Burbo and Rhyl Flats in the UK were given in press releases as $1.9 million $2.2 million respectively. Jacketed foundations are more robust than monopile foundations and are typically used with 5.0 MW turbines and/or in deep water. Two press releases for Nordsee Ost (Germany) and Ormonde (UK) put the unit costs for this type of foundation at $2.74 million and $2.43 million respectively. A 2012 release by the European Energy Programme for Recovery put the cost of deepwater (40 meters) gravity foundations at Global Tech 1 (Germany) as $6.65 million per foundation.

All foundations should feature an increasing cost with depth as material costs will necessarily be higher; however, this is not captured in this model currently due to the paucity of project cost data to estimate such a relationship. Jacquemin et al (2011) used field data to estimate foundation weight as a function of water depth; however the data and functions are not given making it impossible to replicate their work. Nonetheless, this source does provide a means to approximate different foundation technology costs including floating foundation technology. Samoteskul et al (2014) demonstrate how the data from Jacquemin et al (2011) can be used in this way.

Electricity Transmission

Electricity transmission equipment is much harder to model at the component level because the optimal transmission system design varies considerably with local conditions and wind farm design. Depending on the size of the farm and its distance from shore, offshore platforms with voltage transformers, converters, and switchgear may be needed. Additionally, there is a critical point where a wind farm's distance from the grid requires a switch from alternating current (AC) power to direct current (DC) power to overcome line losses which reduce the amount of energy delivered. Given design variation across different contexts, we utilized a top-down modeling approach for transmission costs to allow the model to be used broadly without the need for exhaustive system modeling and unit cost information. We collected information about electricity transmission costs (including installation) from 20 wind farms and used it to estimate a relationship between total costs and farm characteristics. This data was collected from the U.K. Ofgem tender process (http://www.ofgem.gov.uk/Networks/offtrans/Pages/Offshoretransmission.aspx) and is shown in Table 2.

Farm	Cost (2012 $Million)	MW	Depth (m)	DC	Land Cable (km)	Sea Cable (km)	Tot Cable (km)
Barrow	52.73	90	14	0	3	27	30
Robin Rigg	93.25	180	6	0	1.8	12.5	14.3
Gunfleet Sands 1 & 2	75.8	173	6.5	0	3.8	9.3	13.1
Sheringham Shoal	285.08	315	18.5	0	21.5	22.4	43.9
Ormonde	158.27	150	19	0	2.8	43	45.8
Greater Gabbard	495.64	504	20.5	0	0.6	45.5	46.1
Thanet	255.34	300	18.5	0	2.4	26.3	28.7
Walney 1	173.56	183	21	0	2.7	45.3	48
Walney 2	164.38	183	27	0	5	43.7	48.7
Gwynt y Mor	449.38	576	27.5	0	11	41.4	52.4
Lincs	456.44	250	9.5	0	12	48	60
London Array Phase 1	699.28	630	11.5	0	0.8	54	54.8
Nordergrunde	89.82	111	7	0	4	28	32
Dolwin 1	1221.81	800	30.5	1	90	75	165
Dolwin 2	1021.5	900	29	1	90	45	135
Helwin 2	817.51	690	28.5	1	45	85	130
Sylwin 1	1393.38	864	26	1	45	160	205
Helwin 1	718.71	576	23	1	45	85	130
Borwin 2	718.71	800	40	1	75	120	195
Borwin 1	598.83	400	40	1	75	125	200

Table 2: Offshore energy transmission infrastructure.

Using an ordinary least squares regression, we estimated the following equation that relates total transmission costs to farm capacity and total transmission cable distance:

$$TransCost = \beta_0 MW + \beta_1 TotCable + \epsilon \tag{5.88}$$

To capture the effect of transmission losses due to resistance, we estimated this separately for each current type (AC and DC). Since our data suggest a critical threshold of greater than 54.8km for DC transmission, we adopt 60km as the transition point. This is also consistent with published figures regarding the cost effectiveness of transitioning from AC to DC transmission (Carbon Trust, 2008; UMaine, 2011); see Table 3

	Costs if $\leq$ 60km (AC)	Costs if > 60km (DC)
MW	.81***	1.09**
	-0.15	-0.37
Cables (km)	1.36	0.89
	-1.19	-1.61
Adj R^2	0.937	0.951

*Table 3, AC DC transmission costs. *p<.10, **p<.05, ***p<.01*

These results provide a predictive model of transmission costs as a function of current type, total farm capacity in MW, and the total length of transmission cable in km. To calculate the total length of transmission cable from any given offshore location, the model requires some information about the onshore grid. The provided options are meant to provide the user flexibility based on data availability and common analysis questions. The user has two options:

- Create a *.csv* table that includes latitude and longitude details for all grid connection points in the area of interest
- Use a fixed parameter to model grid location

The table option gives the user the ability to indicate both landing points on the coastline and grid connection points. For each potential wind farm site (each ocean pixel that fits the other constraints of the model and is in the AOI), the model identifies the closest specified land point and calculates the straight-line distance to that point. It then finds the closest grid connection point and calculates the straight-line distance to that point. Summing these two distances yields the total length of the transmission cables used in the calculation for transmission costs in Table 3. The user can optionally omit landing points from the table and only include grid points: in this case the model simply calculates total length of the transmission cable as the straightline distance from each potential wind farm location to the nearest grid point.

The fixed parameter option specifies a mean distance inland along the entire coast that represents the expected distance that overland cables may have to travel to reach a grid connection. Since grid connection points for large farms are very opportunistic and represent a relatively small portion of capital costs, it is not unrealistic to model grid connection this way in the absence of a detailed grid connection scheme. The default parameter included, 5.5 km, is the mean overland cable distance from the UK from the transmission infrastructure table above.

Above and beyond the cost of sending the energy to shore, wind farms also require cables which connect turbines to each other, called array cables. We estimated a simple linear relationship between array cables and the number of turbines based on the data given below:

Farm	Location	# of Turbines	km of cable	Total Cost ($mil)
Nordsee Ost	Germany	48	63	14.3
Amrumbank	Germany	80	86	27.1
Gwynt y Mor	UK	160	148	34.7
Anholt	Denmark	111	160	41.4
Baltic 2	Germany	80	80	32.5
Sheringham Shoal	UK	88	88	17.7

Table 4. Array cabling

The data above suggest that .91km of cable is required per turbine at a cost of $260,000 per km. This establishes a relationship of array cable to wind turbines which can retrieve the total cost of array cable based only on the number of turbines in the farm.

Other Costs

There are a variety of additional costs associated with the construction phase, such as those for development, engineering, procurement, and royalties. AWS Truewind (2010) estimate these costs to amount to 2% of total capital expenditures; Blanco (2009) indicates it could be as high as 8%. We adopt their method of using a ratio of capital costs for calculating these costs and use the mean value of 5% as the default .

Installation of foundations, turbines, and transmission gear (cables and substations) comprises its own cost category. Kaiser and Snyder (2012) take a comprehensive view of installation costs and find that installation costs make up approximately 20% of capital expenditures in European offshore wind farms. Accordingly, this model treats installation costs as a fixed percentage of total capital costs and uses the default value suggested by Kaiser and Snyder (2012).

Decommissioning the facility at the end of its useful life ($t = T$) enters into the model in a similar way as installation costs, in that it is a fixed fraction of capital expenditures. Snyder and Kaiser (2012) put this one-time cost at 2.6% to 3.7% of initial expenditures (net of scrap value) for the Cape Wind farm using a sophisticated decommissioning model. The default value used in this model is 3.7%.

Most of the costs of an offshore wind energy farm are related to the initial capital costs; however, there are ongoing costs related to maintenance and operations (O&M) as well. Boccard (2010) uses a methodology consistent with the rest of our modeling by calculating annual O&M cost as a % of original capital costs, and puts the costs somewhere between 3 and 3.5. The default value used in this model is 3.5%, and can be changed along with all the other costs in this section by editing the global parameters *.csv* file.

Energy Prices

This model is currently designed to accept a fixed unit price for a kilowatt hour (kWh) of energy over the lifetime of the wind farm. In some locations, wind farm operators receive a subsidized rate known as a feed-in tariff which guarantees them a set price for their energy over some time horizon. In other locations, wind farm operators must negotiate with energy providers and public utility commissions to secure a power purchase agreement. These are contracts that specify a unit price for energy delivered. We do not supply a default unit price for energy as energy prices fluctuate widely over space and government policies may significantly influences prices in different countries.

Levelized Cost of Energy

The levelized cost of energy (http://en.wikipedia.org/wiki/Cost_of_electricity_by_source) (LCOE) is the unit price that would need to be received for energy that would set the present value of the project equal to zero. As such, it gives the lowest price/kWh that a wind farm developer could receive before they considered a project not worthwhile. The output given by the model is in terms of $/kWh and is calculated as:

$$LCOE = \frac{\sum_{t=1}^{T} \frac{O\&M \cdot CAPEX}{(1+i)^t} + \frac{D \cdot CAPEX}{(1+i)^T} + CAPEX}{\sum_{t=1}^{T} \frac{E_t}{(1+i)^t}}$$

Where $CAPEX$ is the initial capital expenditures, $O\&M$ is the operations and management parameter, D is the decommissioning parameter, E_t is the annual energy produced in kWh, i is the discount or WACC rate, and t is the annual time step, where $t = \{1 \ldots T\}$.

Validation

Capital Cost Model

Since capital expenditures represent the largest proportion of costs, and much of the ancillary costs are fixed fractions of capital costs, it is critically important to validate our model against stated offshore wind farm costs worldwide. To do so, we collected data from http://www.4coffshore.com/ and http://www.lorc.dk/offshore-wind-farms-map/statistics on stated capital costs and designs for wind farms that are in construction or currently operational. We constrained the data collection to only those employing 3.6 MW and 5.0 MW turbines, for which we have provided default data with the InVEST model. Stated capital costs gathered from 4Coffshore were inflated to 2012 $US using their supplied financial close information as the basis for when the cost estimate was collected. To generate predictions, the design of each farm was input into the InVEST model using appropriate default cost parameters for all components. Most farms have their own electrical transmission equipment, though some deepwater farms are beginning to used centralized offshore substations that aggregate energy for transport from multiple farms. To predict electrical transmission costs for these farms, it was first necessary to estimate the cost of the entire offshore substation and then attribute a prorated capital cost to each farm based on their relative contribution to exported energy capacity. For example, an offshore substation with a 800 MW export capacity that is connected to Farm A (200 MW) and Farm B (600 MW) would contribute 25% of capital costs to Farm A and 75% to Farm B. The results of our validation show a very strong correlation between predictions and stated capital costs for 3.6 MW and 5.0 MW turbines using the default data (see Figure 5.6).

Since this model was released in early 2013, it has been tested against other modeling approaches. They are noted below for reference:

1. The InVEST model was compared alongside model estimates from the National Renewable Energy Laboratory (NREL) and a consulting firm in a report out of the University of California, Santa Barbara, that measured the levelized cost of wind energy in Bermuda. InVEST was within 3% of the NREL estimate and 12% of the estimate made by the consulting firm. http://www.bren.ucsb.edu/research/2014Group_Projects/documents/BermudaWind_Final_Report_2014-05-07.pdf

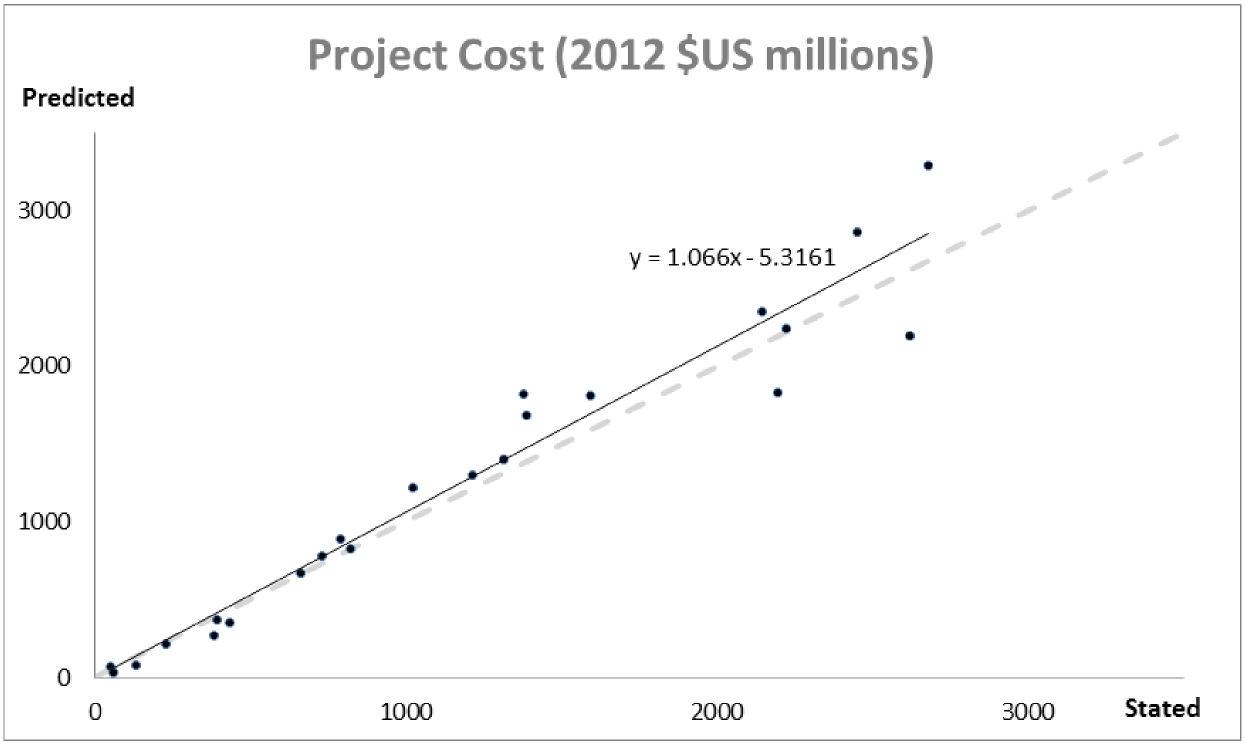


Fig. 5.40: Predicted capital costs versus stated capital costs.

Limitations and Simplifications

Energy Production

The quality of wind input data determines the accuracy of model results. So, users need to understand the quality of wind input data for proper interpretation of the model results. The default wind input data are more appropriate for global and regional scale applications at 4 or 60 minutes spatial resolution.

Harvested wind energy indicates the averaged energy output for a given period based on the output power curve of a wind turbine. Users may want to consider additional technology-specific information, such as device availability, power conversion efficiency, and directional factors by applying adjustment factors to the harvested energy output.

Energy Valuation

As the validation section demonstrates, the model and the default data reliably predict capital costs using the supplied inputs. Revenues are linked to energy production and a user-entered price. Currently the model is not capable of handling a price schedule with time-variant energy pricing. Inflation can be incorporated into the discount rate parameter, but more elaborate price changes are currently outside of the model's capability.

More reliable cost projections over space could likely be attained by:

- Creating a foundation cost function that accounts for higher costs in deeper waters
- Having installation costs vary as a function of bottom geology

These are features that are being explored for subsequent model updates.

The model is amenable to producing valuation outputs for floating turbines, but was not designed specifically for this task. To produce outputs, the user needs to input reasonable values for depth and distance constraints as well as "foundation" costs equal to the unit cost of the aggregate equipment needed to float a turbine. The electrical transmission model was derived using technologies that are suitable to roughly 60 meters depth and 200 kilometers distance from shore and will likely produce less accurate cost projections outside of those bounds.

5.12.4 Data Needs

Required Inputs

1. **Workspace (required).** Select a folder to be used as your workspace. If the folder you select does not exist, a new one will be created. This folder will contain the rasters produced by this model. If datasets already exist in this folder, they will be overwritten. The output will be contained in an folder named *output* inside the workspace directory.
2. **Wind Data Points (required).** A binary file that represents the wind input data (Weibull parameters). This parameter box should point to one of two files provided by our model. These files are found in the *WindEnergy-input* direction inside the InVEST installation directory.

- Global Data: *GLobal_EEZ_WEBPAR_90pct_100ms.bin*
- East Coast of the US: *ECNA_EEZ_WEBPAR_Aug27_2012.bin* for finer resolution of that area.

3. **Area Of Interest (AOI)** An optional polygon shapefile that defines the area of interest. The AOI must be projected with linear units equal to meters. If the AOI is provided it will clip and project the outputs to that of the AOI. The Distance inputs are dependent on the AOI and will only be accessible if the AOI is selected. If the AOI is selected and the Distance parameters are selected, then the AOI should also cover a portion of the land polygon to calculate distances correctly. An AOI is required for valuation.
4. **Bathymetry (DEM)** A raster dataset for the elevation values in meters of the area of interest. The DEM should cover at least the entire span of the area of interest and if no AOI is provided then the default global DEM should be used.
5. **Land Polygon for Distance Calculation** A polygon shapefile that represents the land and coastline that is of interest. For this input to be selectable the AOI must be selected. The AOI should also cover a portion of this land polygon to properly calculate distances. This coastal polygon, and the area covered by the AOI, form the basis for distance calculations for wind farm electrical transmission. This input is required for masking by distance values and for valuation.
6. **Global Wind Energy Parameters** A CSV file that holds wind energy model parameters for both the biophysical and valuation modules. These parameters are defaulted to values that are reviewed in the **The Model** section of this guide. We recommend careful consideration before changing these values.
7. **Results Suffix** A String that will be added to the end of the output file paths.
8. **Turbine Type** A CSV file that contains parameters corresponding to a specific turbine type. The InVEST package comes with two turbine model options, 3.6 MW and 5.0 MW. You may create a new turbine class (or modifying existing classes) by using the existing file format conventions and filling in your own parameters. It is recommended that you do not overwrite the existing default CSV files. These files are found in the *WindEnergyinput* direction inside the InVEST installation directory and named

- 3.6 MW: *3_6_turbine.csv*
- 5.0 MW: *5_0_turbine.csv*

9. **Minimum Depth for Offshore Wind Farm Installation (m)** A floating point value in meters for the minimum depth of the offshore wind farm installation.
10. **Maximum Depth for Offshore Wind Farm Installation (m)** A floating point value in meters for the maximum depth of the offshore wind farm installation.

11. **Minimum Distance for Offshore Wind Farm Installation (m)** A floating point value in meters that represents the minimum distance from shore for offshore wind farm installation. Required for valuation.
12. **Maximum Distance for Offshore Wind Farm Installation (m)** A floating point value in meters that represents the maximum distance from shore for offshore wind farm installation. Required for valuation.

Valuation

13. **Cost of the Foundation Type (millions of dollars)** A floating point number for the unit cost of the foundation type (in millions of dollars). The cost of a foundation will depend on the type of foundation selected, which itself depends on a variety of factors including depth and turbine choice.
14. **Number Of Turbines** An integer value indicating the number of wind turbines per wind farm.
15. **Price of Energy per Kilowatt Hour ($/kWh)** The price of energy per kilowatt hour.
16. **Discount Rate** The discount rate reflects preferences for immediate benefits over future benefits. Enter in decimal form (Ex: 1% as 0.01, 100% as 1.0).
17. **Grid Connection Points** An optional CSV file with grid and land points to determine energy transmission cable distances from. Each point location is represented as a single row with columns being *ID*, *TYPE*, *LATI*, and *LONG*. The *LATI* and *LONG* columns indicate the coordinates for the point. The *TYPE* column relates to whether it is a land or grid point. The *ID* column is a simple unique integer. The shortest distance between respective points is used for calculations. An example:

ID	TYPE	LATI	LONG
1	GRID	42.957	-70.786
2	LAND	42.632	-71.143
3	LAND	41.839	-70.394

18. **Average Shore to Grid Distance (km)** A number in kilometers that is only used if grid points are NOT used in valuation. When running valuation using the land polygon to compute distances, the model uses an average distance to the onshore grid from coastal cable landing points instead of specific grid connection points.

5.12.5 Interpreting Results

All output resolutions are based on the resolution of the supplied digital elevation model raster. When the resolution of the DEM exceeds the resolution of the wind data layers, pixel values are determined by using bilinear interpolation.

- *carbon_emissions_tons.tif*: a GeoTIFF raster file that represents tons of offset carbon emissions for a farm built centered on a pixel per year.
- *density_W_per_m2.tif*: a GeoTIFF raster file that represents power density (W/m^2) centered on a pixel.
- *example_size_and_orientation_of_a_possible_wind_farm.shp*: an ESRI shapefile that represents the outer boundary of a sample windfarm. The position of this polygon is random and is meant to give the user a sense of scale of the potential wind farm.
- *harvested_energy_MWhr_per_yr.tif*: a GeoTIFF raster file that represents the annual harvested energy from a farm centered on that pixel.
- *levelized_cost_price_per_kWh.tif*: a GeoTIFF raster file that represents the unit price of energy that would be required to set the present value of the farm centered at that pixel equal to zero.
- *npv_US_millions.tif*: a GeoTIFF raster file that represents the net present value of a farm centered on that pixel.
- *wind_energy_points.shp*: an ESRI Shapefile that summarizes the above outputs for each point...

5.12.6 Data Sources

Energy Output Data

- Wind time series data: NOAA's National Weather Service provides hindcast reanalysis results for wind time series; http://polar.ncep.noaa.gov/waves/index2.shtml. The spatial resolution of the model results ranges from 4 to 60 minutes depending on the global and regional grid systems. The model outputs have been saved at 3-hour interval from 1999 to the present. The model results have been validated with ocean buoy data at many locations and provide good quality wind information.
- Water depth: NOAA's National Geophysical Data Center (NGDC) provides global bathymetry data with various spatial resolutions at http://www.ngdc.noaa.gov/mgg/bathymetry/relief.html.
  - ETOPO1 is a 1 arc-minute global relief model of Earth's surface that integrates land topography and ocean bathymetry. It was built from numerous global and regional data sets, and is available in "Ice Surface" (top of Antarctic and Greenland ice sheets) and "Bedrock" (base of the ice sheets) versions. NGDC also provides regional and other global bathymetry datasets.
- LORC knowledge provides the parameter information of offshore wind turbines that are currently operating in the world. http://www.lorc.dk/offshore-wind-farms-map/list?sortby=InstalledCapacity&sortby2=&sortorder=desc

Valuation

Data sources are largely cited above, except for figures that were derived from press releases. Press releases were found by an exhaustive Google keyword search on "offshore wind energy" contract and several variants of that theme. All costs were recorded and inflated in their original currency and exchanged to $US at the spot rate on March 30th, 2012.

This file (http://ncp-dev.stanford.edu/~dataportal/Wind_Sources.zip) contains an archive of the sources sited for costs and a spreadsheet that links each cost figure to the relevant press release, conference proceeding, etc.

5.12.7 Running the Model

To run the wind energy model, navigate to the "Wind Energy" application under the windows Start Menu found in *All Programs->InVEST{version}*. The user interface will indicate the required and optional input arguments as described in the **Data Needs** section above. Click the *Run* button to start the model. A successful run will be indicated in the window and a file explorer will open containing the results.

If you encounter any issues please post to the user's support forum at http://ncp-yamato.stanford.edu/natcapforums.

5.12.8 References

AWS Truewind. 2010. New York's Offshore Wind Energy Development Potential in the Great Lakes. Feasibility Study for New York State Energy Research and Development Authority.

Blanco, M. 2009. The Economics of Wind Energy. Renewable and Sustainable Energy Reviews, 13, 1372-82. http://dx.doi.org/10.1016/j.rser.2008.09.004

Boccard, N. 2010. Economic Properties of Wind Power: A European Assessment. Energy Policy, 38, 3232-3244. http://dx.doi.org/10.1016/j.enpol.2009.07.033

Carbon Trust. 2008. Offshore Wind Power: Big Challenge, Big Opportunity. Report on behalf of the Government of the United Kingdom.

Celik, A. N. 2003. A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern of Turkey. Renewable Energy 29:509-604. http://dx.doi.org/10.1016/j.renene.2003.07.002

Elliott, D. L., C. G. Holladay, W. R. Barchet, H. P. Foote, and W. F. Sandusky. 1986. Wind energy resource atlas of the United States. DOE/CH 10093-4. Solar Technical Information Program, Richland, Washington.

Global Wind Energy Council (GWEC). 2013. Global Wind Statistics, 2012. Accessed at: http://www.gwec.net/wp-content/uploads/2013/02/GWEC-PRstats-2012_english.pdf

Griffin, R., Buck, B., and Krause, G. 2015a. Private incentives for the emergence of co-production of offshore wind energy and mussel aquaculture. Aquaculture, 346, 80-89. http://dx.doi.org/10.1016/j.aquaculture.2014.10.035

Griffin, R., Chaumont, N., Denu, D., Guerry, A., Kim, C., and Ruckelshaus, M. 2015b. Incorporating the visibility of coastal energy infrastructure into multi-criteria siting decisions. Marine Policy, 62, 218-223. http://dx.doi.org/10.1016/j.marpol.2015.09.024

Hsu, S. A., E. A. Meindl, and D. B. Gilhousen. 1994. Determining the power-law wind-profile exponent under near-neutral stability conditions at sea. Journal of applied meteorology 33:757-765. http://dx.doi.org/10.1175/1520-0450(1994)033%3C0757:DTPLWP%3E2.0.CO;2

Jacquemin, J., Butterworth, D., Garret, C., Baldock, N., and A. Henderson. 2011. Inventory of location specific wind energy cost. WP2 Report D2.2. Spatial deployment of offshore wind energy in Europe (Wind-Speed). Garrad Hassan & Partners Ltd. Supported by Intelligent Energy Europe. Available at: http://www.windspeed.eu/media/publications/WINDSPEED_D2_2_revised_May_2011.pdf

Jafarian, M., and A. M. Ranjbar. 2010. Fuzzy modeling techniques and artificial neural networks to estimate annual energy output of a wind turbine. Renewable Energy 35:2008-2014. http://dx.doi.org/10.1016/j.renene.2010.02.001

Kaiser, M. and B. Snyder. 2012. Offshore wind capital cost estimation in the U.S. Outer Continental Shelf: A reference class approach. Marine Policy, 36, 1112-1122. http://dx.doi.org/10.1016/j.marpol.2012.02.001

Levitt, A., Kempton, W., Smith, A., Musial, W., and J. Firestone. 2011. Pricing offshore wind energy. Energy Policy, 39, 6408-6421. http://dx.doi.org/10.1016/j.enpol.2011.07.044

Lorc Knowledge. 2012. List of offshore wind farms. http://www.lorc.dk/offshore-wind-farms-map/list Accessed at December 31, 2012.

Manwell, J. F., J. G. Mcgowan, and A. L. Rogers. 2009. Wind energy explained: Theory, design and application. John Wiley & Sons Ltd., West Sussex, United Kingdom.

National Oceanic and Atmospheric Administration. 1976. U. S. Standard Atmosphere. NOAA- S/T76-1562, Washington, DC.

Pallabazzer, R. 2003. Provisional estimation of the energy output of wind generators. Renewable Energy, 29, 413-420. http://dx.doi.org/10.1016/S0960-1481(03)00197-6

Samoteskul, K., Firestone, J., Corbett, J., and J. Callahan. 2014. Changing vessel routes could significantly reduce the cost of future offshore wind projects. Journal of Environmental Management, 141, 146-154. http://dx.doi.org/10.1016/j.jenvman.2014.03.026

UMaine. 2011. Maine deepwater offshore wind report. http://www.deepcwind.org/docs/OfficialOffshoreWindReport-22311.pdf

5.13 Marine Finfish Aquacultural Production

5.13.1 Summary

Supporting the production of aquacultured fish and shellfish is an important service provided by coastal and marine environments. Because salmon is one of the two most important finfish in aquaculture worldwide, the current version of the InVEST aquaculture model analyzes the volume and economic value of Atlantic salmon (*Salmo salar*) grown in netpen aquaculture facilities based on farming practices, water temperature, and economic factors. Inputs for the present model include farm location, management practices at the facilities, water temperature, economic data for valuation, and the time period over which results are of interest. This model is best used to evaluate how human activities (e.g., the addition or removal of farms or changes in harvest management practices) and climate change (e.g., change in sea surface temperature) may affect the production and economic value of aquacultured Atlantic salmon. Limitations of the model include assumptions that harvest practices, prices, and costs of production of aquacultured fish are constant for the selected time period. Additionally, risk of disease outbreaks and variability between individual salmon within a farm are not included in the model. Future releases of this model will include the following features: 1) guidance for modifying the Atlantic salmon model for other aquacultured marine fish, 2) quantification of wastes produced at aquaculture facilities, 3) a separate module for quantifying volume, economic value, filtration and production of wastes of aquacultured shellfish (e.g., oyster, shrimp), and 4) a sub-module to evaluate impacts of parasitic sea-lice on farmed Atlantic salmon.

5.13.2 Introduction

Human demand for protein from the ocean has rapidly increased and is projected to continue to do so in coming decades (Delgado et al. 2003, Halwart et al. 2007, Soto et al. 2008). In recent years, the scales, previously tilted towards provisioning of protein from capture fisheries, have shifted toward aquaculture. In particular, finfish aquaculture, primarily for Atlantic salmon, has intensified in coastal areas over the past two decades (FAO 2004, Goldburg and Naylor 2004, Naylor and Burke 2005). In 2002, farmed salmon production, over 90% of which was for Atlantic salmon, was 68% higher than the volume of wild capture (FAO 2004). Atlantic salmon farming, conducted in floating netpens in low energy, nearshore areas, is a well-established, consolidated industry that operates in the temperate waters of Norway, Chile, the United Kingdom and Canada.

Commercial operations for Atlantic salmon use the marine environment to produce a valuable commodity, which generates revenue and is a source of employment. Yet salmon farming is controversial due to potentially adverse impacts to marine ecosystems and, thereby, people who derive their livelihoods from those ecosystems (e.g., commercial fishermen, tourism operators). Concerns about the effects of Atlantic salmon aquaculture on the marine ecosystem involve debate about the impacts of emission of dissolved and solid wastes to water quality and living habitats, degradation of water quality due to use of antibiotics, mixing and competition of escaped farmed salmon with endemic species (e.g., Pacific salmon), increased risk of parasitism and disease, and depletion of forage fish resources harvested from other ecosystems for use as Atlantic salmon feed.

Regulations for the Atlantic salmon aquaculture industry vary regionally, from the most stringent requirements for locating and operating facilities in Norwegian waters, to fewer constraints for farms in Chilean waters. For all operations, there are regulatory limits on where and how aquaculture can be conducted and requirements for monitoring and regulating the amount of waste generated at different facilities, and in some cases, mitigation requirements.

Weighing the economic benefits of Atlantic salmon aquaculture against the environmental costs involves quantifying both. The InVEST model presented here does the former by quantifying the volume and economic value of the commodity. Future outputs will include dissolved and particulate wastes generated as a byproduct of Atlantic salmon production. These outputs will be available for use in other InVEST models (e.g., water quality, habitat quality, fisheries) to assess impacts of Atlantic salmon aquaculture on other coastal and marine ecosystem services. With the full suite of model outputs, InVEST users will be able explore how different spatial configuration of Atlantic salmon farms in their region affects other ecosystem benefits and alleviates or exacerbates tradeoffs between economic benefits and downstream environmental costs.

5.13.3 The Model

The model is designed to address how the production and economic value of farmed Atlantic salmon at individual aquaculture facilities and across a user-defined study area change depending on farm operations and changes in water temperature. Temporal shifts in price, costs or harvest management practices are not dynamically modeled, but can be represented by running the model sequentially, where each run uses different information on prices, costs and farm operations. The risk of disease outbreaks and variability between individual salmon within a farm are not included in the model. The model will yield the most accurate outputs when parameterized with site-specific temperature and farm operations data. If site-specific data are unavailable, the provided ranges of default values can be used to yield first approximations of results (see *Data Needs* section).

The model is run simultaneously for all Atlantic salmon farms identified by the user. Each farm can have a user-defined set of operations and management practices. The volume of fish produced on a farm depends on water temperature (which affects growth), the number of fish on the farm, the target harvest weight range, and the mortality rate. Fish growth is modeled on a daily time-step until the fish reach the target harvest weight range, after which they are harvested. After a user-defined fallowing period, the farm is restocked and this initiates the next production cycle. Production cycles continue for each farm until the end of the time period of interest (e.g., 2 years, 10 years). Outputs include the harvested weight of fish and net revenue per cycle for each individual farm. In addition, the model yields a map of the total harvested weight, total net revenue, and net present value over the time period of interest.

How it Works

The model runs on a vector GIS dataset that maps individual aquaculture facilities for Atlantic salmon that are actively farmed over a user-defined time period. The map can be based on current farming (the "status quo" or "baseline" scenario), or on scenarios of projected expansion or contraction of the industry or on projected changes in water temperature.

In each farm we model the production of fish in three steps. (1) We model the growth of individual fish to harvest weight. (2) We calculate the total weight of fish produced in each farm as the number of fish remaining at harvest, multiplied by their harvested weight, less the weight removed during processing (gutting, etc.) and the weight of fish lost to natural mortality. (3) Lastly, all the fish in a farm are harvested at the same time, and the farm is restocked after a user-defined fallowing period. Valuation of processed harvest is an optional fourth step in the model.

Growth of the Individual Fish to Harvest Weight

Atlantic salmon weight (kg) is modeled from size at outplanting to target harvest weight. Weight is a function of growth rate and temperature (Stigebrandt 1999). Outplanting occurs when Atlantic salmon have been reared beyond their freshwater life stages. The model runs on a daily time step because the next version of the model (to be released in May 2011) will quantify waste products from aquaculture farms for use as inputs into the Marine InVEST water quality model. Fine resolution temporal data are more appropriate for the seasonal evaluation of environmental impacts (e.g., seasonal eutrophication).

Weight W_t at time t (day), in year y, and on farm f is modeled as:

$$W_{t,y,f} = (\alpha W^{\beta}_{t-1,y,f} \cdot e^{T_{t-1,f}\tau}) + W_{t-1,y,f} \tag{5.89}$$

where α ($g^{1-b}day^{-1}$) and β (non-dimensional) are growth parameters, $T_{t,f}$ is daily water temperature (C) at farm f, and τ (C^{-1}) represents the change in biochemical rates in fishes with an increase in water temperature. The value for Atlantic salmon (0.08) indicates a doubling in growth with an 8-9 C increase in temperature. Daily water temperatures can be interpolated from monthly or seasonal temperatures. The growing cycle for each farm begins on the user-defined date of outplanting ($t = 0$). The outplanting date is used to index where in the temperature time series to begin. The initial weight of the outplanted fish for each farm is user-defined. An individual Atlantic salmon grows until it reaches its target harvest weight range, which is defined by the user as a target harvest weight.

Total Weight of Fish Produced per Farm

To calculate the total weight of fish produced for each farm, we assume that all fish on a farm are homogenous and ignore variability in individual fish growth. This assumption, though of course incorrect, is not likely to affect the results significantly because 1) netpens are stocked so as to avoid effects of density dependence and 2) aquaculturists outplant fish of the same weight to netpens for ease of feeding and processing. We also assume that when fish reach a certain size, all fish on the farm are harvested. In practice, farms consist of several individual netpens, which may or may not be harvested simultaneously. If a user has information about how outplanting dates and harvest practices vary between netpens on a farm, the user can define each netpen as an individual "farm."

The total weight of processed fish TPW on farm f in harvest cycle c:

$$TPW_{f,c} = W_{t_h,h,f} \cdot d \cdot n_f e^{-M \cdot (t_h - t_0)} \tag{5.90}$$

where $W_{t_h,h,f}$ is the weight at date of harvest t_h, y on farm f from Equation (5.89), d is the processing scalar which is the fraction of the fish in the farm that remains after processing (e.g., weight of headed/gutted or filleted fish relative to harvest weight), n_f is the user-defined number of fish on farm f, and $e^{-M \cdot (t_h - t_o)}$ is the daily natural mortality rate M experienced on the farm from the date of outplanting (t_0) to date of harvest (t_h).

Restocking

The previous 2 steps describe how fish growth is mdoeled for one production cycle. However, the user may want to evaluate production of fish over a series of production cycles. The primary decision to be made when modeling multiple harvest cycles is if (and if so, how long) a farm will be left to lie fallow after harvest and before the next production cycle begins (initiated by outplanting).

If used, fallowing periods are considered hard constraints in the model such that a farm cannot be restocked with fish until it has lain fallow for the user-defined number of days. This is because fallowing periods are often used to meet regulatory requirements, which can be tied to permitting, and thus provide incentive for compliance. Once fish are harvested from a farm and after the user-defined fallowing period, new fish are outplanted to the farm. The model estimates the harvested weight of Atlantic salmon for each farm in each production cycle. The total harvested weight for each farm over the time span of the entire model run is the sum of the harvested weights for each production cycle.

Valuation of Processed Fish (Optional)

The aquaculture model also estimates the value of that harvest for each farm in terms of net revenue and net present value (NPV) of the harvest in each cycle. The net revenue is the harvest weight for each cycle multiplied by market price, where costs are accounted for as a fraction of the market price for the processed fish. Fixed and variable costs, including costs of freshwater rearing, feed, and processing will be more explicitly accounted for in the next iteration of this model. The NPV of the processed fish on a farm in a given cycle is the discounted net revenue such that:

$$NPV_{f,c} = TPW_{f,c}[p(1-C)] \cdot \frac{1}{(1+r)^t} \tag{5.91}$$

where $TPW_{f,c}$ is the total weight of processed fish on farm f in harvest cycle c, p is the market price per unit weight of processed fish, C is the fraction of p that is attributable to costs, r [4] is the daily market discount rate, and t is the number of days since the beginning of the model run.

Note: The beginning of the model run is the initial outplanting date for the very first farm (of all the farms in the study area) to receive fish. Thus, the net revenue for each farm in each harvest cycle is discounted by the number of days since the very first farm was initially stocked. The total NPV for each farm over the duration of the model run is the discounted net revenue from each harvest cycle summed over all harvest cycles c.

The discount rate reflects society's preference for immediate benefits over future benefits (e.g., would you rather receive \$10 today or \$10 five years from now?). The default annual discount rate is 7% per year, which is one of the rates recommended by the U.S. government for evaluation of environmental projects (the other is 3%). However, this rate can be set to reflect local conditions or can be set to 0%.

[4] The daily discount rate is computed as the annual discount rate divided by 365. For an annual discount rate of 7%, the daily discount rate is 0.00019178.

Uncertainty Analysis (Optional)

Optionally, if the fish growth parameters are not known with certainty, the model can perform uncertainty analysis. This uncertainty analysis is done via a Monte Carlo simulation. In this simulation, the growth parameters (α and b) are repeatedly sampled from a given normal distribution, and the model is run for each random sampling.

The results for each run of the simulation (harvested weight, net present value, and number of completed cycles per farm) are collected and then analyzed. Uncertainty results are output in two ways: first, the model outputs numerical results, displaying the mean and the standard deviation for all results across all runs. Second, the model creates histograms to help visualize the relative probability of different outcomes.

5.13.4 Limitations and Simplifications

Limitations of the model include assumptions that harvest practices, prices, and costs of production of aquacultured fish are constant over the selected time period. Additionally, risk of disease outbreaks and variability between individual salmon within a farm are not included in the model.

The current model operates at a daily time step (requiring daily temperature data).

Uncertainty in input data is currently supported only for fish growth parameters. There is currently no support for uncertainty in input data such as water temperature.

Growth is assumed to be exponential up to the point of harvesting. Survival and growth do not depend on density. The assumption is that aquaculturists are optimizing the stocking density such that there is not excess mortality due to over-crowding.

5.13.5 Data Needs

Data Sources

Here we outline the specific data and inputs used by the model and identify potential data sources and default values. Four data layers are required, and one is optional (but required for valuation).

1. **Workspace Location (required).** Users are required to specify a workspace folder path. It is recommended that the user create a new folder for each run of the model. For example, by creating a folder called "runBC" within the "Aquaculture" folder, the model will create "intermediate" and "output" folders within this "runBC" workspace. The "intermediate" folder will compartmentalize data from intermediate processes. The model's final outputs will be stored in the "output" folder.:

```
Name: Path to a workspace folder.  Avoid spaces.
Sample path: \InVEST\Aquaculture\runBC
```

2. **Finfish Farm Location (required).** A GIS polygon or point dataset, with a latitude and longitude value and a numerical identifier for each farm.:

```
Names: File can be named anything, but no spaces in the name
File type: polygon shapefile or .gdb
Rows: each row is a specific netpen or entire aquaculture farm
Columns: columns contain attributes about each netpen (area, location, etc.).
Sample data set: \InVEST\Aquaculture\Input\Finfish_Netpens.shp
```

Note: The user must ensure that one field contains unique integers. This field name (i.e. "FarmID" in the sample data) must be chosen by the user for input #3 as the "farm identifier name".

Note: The model checks to ensure that the finfish farm location shapefile is projected in meters. If it is not, the user must re-project it before running the model.

3. **Farm Identifier Name (required).** The name of a column heading used to identify each farm and link the spatial information from the GIS features (input #2) to subsequent table input data (farm operation and daily water temperature at farm tables, inputs # 6-7). Additionally, the numbers underneath this farm identifier name must be unique integers for all the inputs (#2, 6, & 7).:

```
Names: A string of text identifying a column in the Finfish Farm Location shapefile's attribute
File type: Drop-down option
Sample: FarmID
```

4. **Fish growth parameters (required, defaults provided).** Default a (0.038 g/day), b (0.6667 dimensionless units), and τ (0.08 C^{-1}) are included for Atlantic salmon, but can be adjusted by the user as needed. If the user chooses to adjust these parameters, we recommend using them in the simple growth model (Equation (5.89)) to determine if the time taken for a fish to reach a target harvest weight typical for the region of interest is accurate.:

```
Names: A numeric text string (floating point number)
File type: text string (direct input to the ArcGIS interface)
Sample (default): 0.038 for a / 0.6667 for b
```

5. **Uncertainty analysis data (optional).** These parameters are required only if uncertainty analysis is desired. Users must provide three numbers directly through the tool interface.:

- Standard deviation for fish growth parameter a. This represents uncertainty in the estimate for the value of a.
- Standard deviation for fish growth parameter b. This represents uncertainty in the estimate for the value of b.
- Number of Monte Carlo simulation runs. This controls the number of times that the parameters are sampled and the model is run, as part of a Monte Carlo simulation. A larger number will increase the reliability of results, but will also increase the running time of the model. Monte Carlo simulations typically involve about 1000 runs.

6. **Daily Water Temperature at Farm Table (required).** Users must provide a time series of daily water temperature (C) for each farm in data input #1. When daily temperatures are not available, users can interpolate seasonal or monthly temperatures to a daily resolution. Water temperatures collected at existing aquaculture facilities are preferable, but if unavailable, users can consult online sources such as NOAA's 4 km AVHRR Pathfinder Data and Canada's Department of Fisheries and Oceans Oceanographic Database. The most appropriate temperatures to use are those from the upper portion of the water column, which are the temperatures experienced by the fish in the netpens.:

```
Table Names: File can be named anything, but no spaces in the name
File type: *.xls or .xlsx (if user has MS Office 2007 or newer)
Rows: There are 365 rows (rows 6-370), each corresponding to a day of the year.
Columns: The first two columns contain the number for that year (1-365) and day-month.
Sample: \InVEST\Aquaculture\Input\Temp_Daily.xls\WCVI$
```

Note: For clarification on rows, please refer to the sample temperature dataset in the InVEST package (Temp_Daily.xls).

Note: Column "C" and then all others to its right contain daily temperature data for a specific farm, where the numbers found in row 5 must correspond to the numbers underneath the farm identifier name found in input #2's attribute table.

7. **Farm Operations Table (required).** A table of general and farm-specific operations parameters. Please refer to the sample data table for reference to ensure correct incorporation of data in the model. If you would like to use your own dataset, you can modify values for farm operations (applied to all farms) and/or add new farms (beginning with row 32). However, do not modify the location of cells in this template. If for example, you choose to run the model for three farms only, they should be listed in rows 10, 11 and 12 (farms 1, 2, and 3, respectively). Several default values that are applicable to Atlantic salmon farming in British Columbia are also

included in the sample data table. The majority of these values can be found by talking to aquaculturists in the study area or through regional industry reports from major aquaculture companies (e.g. Panfish, Fjord Seafood, Cermaq, Marine Harvest, Mainstream Canada, and Grieg).

The **General Operation Parameters** of the input table includes the following inputs that apply to all farms: + Fraction of the fish weight (in the farm) remaining after processing (e.g., weight of headed/gutted fish relative to harvest weight) + Natural mortality rate on the farm (daily) + Duration of simulation (in years)

The **Farm-Specific Operation Parameters** of the input table includes the following inputs:

- Rows: Each row in this table (table begins at row #10) contains the input data for a specific farm.
- Columns: Each column contains values and should be named as follows:
  - Farm #: a series of consecutive integers (beginning with "1" in row 10) that identifies each farm and must correspond to the unique integers underneath the farm identifier name found in input #2's attribute table.
  - Weight of fish at start (kg): this is the weight of fish when they are outplanted, which occurs when Atlantic salmon have been reared beyond their freshwater life stages.
  - Target weight of fish at harvest (kg)
  - Number of fish in farm (absolute)
  - Start day for growing (Julian day of the year): this is the date of the initial outplanting at the start of the model run. Outplanting date will differ in subsequent cycles depending on lengths of growth and fallowing periods.
  - Length of fallowing period (number of days): if there is no fallowing period, set the values in this column to "0".

```
Table Names: File can be named anything, but no spaces in the name
File type: *.xls or .xlsx (if user has MS Office 2007 or newer)
Sample: \InVEST\Aquaculture\Input\Farm_Operations.xls\WCVI$
```

8. **Run Valuation? (optional).** By checking this box, users request valuation analysis.
9. **Valuation parameters (required for valuation, defaults provided).**:

```
Names: A numeric text string (positive integer or floating point number)
File type:  text string (direct input to the ArcGIS interface)
Sample (default):
  a. Market price per kilogram of processed fish.
     Default value is 2.25 $/kilogram
     (Urner-Berry monthly fresh sheet reports on price of farmed Atlantic salmon)
  b. Fraction of market price that accounts for costs rather than profit.
     Default value is 0.3 (30%).
  c. Daily market discount rate.
     We use a 7% annual discount rate, adjusted to a daily rate of 0.000192 for 0.0192% (7%/365
```

Note: If you change the market price per kilogram, you should also change the fraction of market price that accounts for costs to reflect costs in your particular system.

5.13.6 Running the Model

The model is available as a standalone application accessible from the Windows start menu. For Windows 7 or earlier, this can be found under *All Programs -> InVEST +VERSION+ -> Finfish Aquaculture*. Windows 8 users can find the application by pressing the windows start key and typing "finfish" to refine the list of applications. The standalone can also be found directly in the InVEST install directory under the subdirectory *invest-3_x86/invest_pollination.exe*.

Viewing Output from the Model

Upon successful completion of the model, a file explorer window will open to the output workspace specified in the model run. This directory contains an *output* folder holding files generated by this model. Those files can be viewed in any GIS tool such as ArcGIS, or QGIS. These files are described below in Section *Interpreting Results*.

5.13.7 Interpreting Results

Model Outputs

The following is a short description of each of the outputs from the Aquaculture tool. Each of these output files is automatically saved in the "Output" folder that is saved within the user-specified workspace directory:

Final results are found in the output folder of the workspace for this model. The model produces two main output files:

- **Output\Finfish_Harvest.shp:** Feature class (copy of input 2) containing three additional fields (columns) of attribute data.
  - Tot_Cycles – The number of harvest cycles each farm completed over the course of the simulation (duration in years)
  - Hrvwght_kg – Total processed weight (in kg, Eqn. 2,) for each farm summed over the time period modeled
  - NPV_USD_1k – The discounted net revenue from each harvest cycle summed over all harvest cycles (in thousands of $). This value will be a "0" if you did not run the valuation analysis.
- **Output\HarvestResults_[date and time].html:** An HTML document containing tables that summarize the inputs and outputs of the model.
  - **Farm Operations** – a summary of the user-provided input data including: Farm ID Number, Weight of fish at start, Weight of fish at harvest, Number of fish in farm, start day for growing and Length of fallowing period
  - **Farm Harvesting** – a summary table of each harvest cycle for each farm including: Farm ID Number, Cycle Number, Days Since Outplanting Date, Harvested Weight, Net Revenue, Net Present Value, Outplant Day, Year
  - **Farm Result Totals** – a summary table of model outputs for each farm including: Farm ID Number, Net Present Value, Number of completed harvest cycles, Total volume harvested
  - **Uncertainty Analysis Results** – this section will be included only if uncertainty analysis was performed. It includes two parts:
    - Numerical Results – a table summarizing mean and standard deviation for model outputs such as harvested weight, net present value, and number of completed harvest cycles.
    - Histograms – a series of histograms to help visualize relative probabilities of different outcomes.

Parameter Log

- Each time the model is run a text file will appear in the workspace folder. The file will list the parameter values for that run and be named according to the date and time.

5.13.8 References

Delgado, C., N. Wada, M. Rosegrant, S. Meijer, and M. Ahmed. 2003. Outlook for Fish to 2020: Meeting Global Demand. Washington, DC: Int. Food Policy Res. Inst.

Farm Operations (input)

Farm ID Number	Weight of Fish at Start (kg)	Weight of Fish at Harvest (kg)	Number of Fish in Farm	Start Day for Growing (1-365)	Length of Fallowing Period (days)
1	0.06	5.4	600000	60	0
2	0.06	5.4	600000	60	0
3	0.06	5.4	590000	50	90
4	0.06	5.4	730000	60	90

Fig. 5.41: First few rows of a sample Farm Operations table in HTML output

Farm Harvesting (output)

Farm ID Number	Cycle Number	Days Since Outplanting Date (Including Fallowing Period)	Length of Given Cycle	Harvested Weight After Processing (kg/cycle)	Net Revenue (Thousands of $)	Net Present Value (Thousands of $)	Outplant Day (Julian Day)	Outplant Year
1	1	512	452	2597703.28318	4091.38267101	3708.35510461	60	1
1	2	1240	453	2599987.46964	4094.98026468	3227.49001717	57	3
2	1	512	452	2597703.28318	4091.38267101	3708.35510461	60	1
2	2	1240	453	2599987.46964	4094.98026468	3227.49001717	57	3
3	1	486	436	2551950.95991	4019.32276186	3661.27108478	50	1
3	2	1214	437	2557467.14914	4028.01075989	3190.59372354	47	3

Fig. 5.42: First few rows of a sample Farm Harvesting table in HTML output

Farm Result Totals (output)

All values in the following table were also populated in the attribute table of the netpens feature class.

Farm ID Number	Net Present Value (Thousands of $) (For Duration of Model Run)	Number of Completed Harvest Cycles	Total Volume Harvested (kg)(After Processing Occurs)
1	6935.8451	2	5197690.7528
2	6935.8451	2	5197690.7528
3	6851.8648	2	5109418.109
4	8498.7201	2	6333350.2645

Fig. 5.43: First few rows of a sample Farm Result Totals table in HTML output

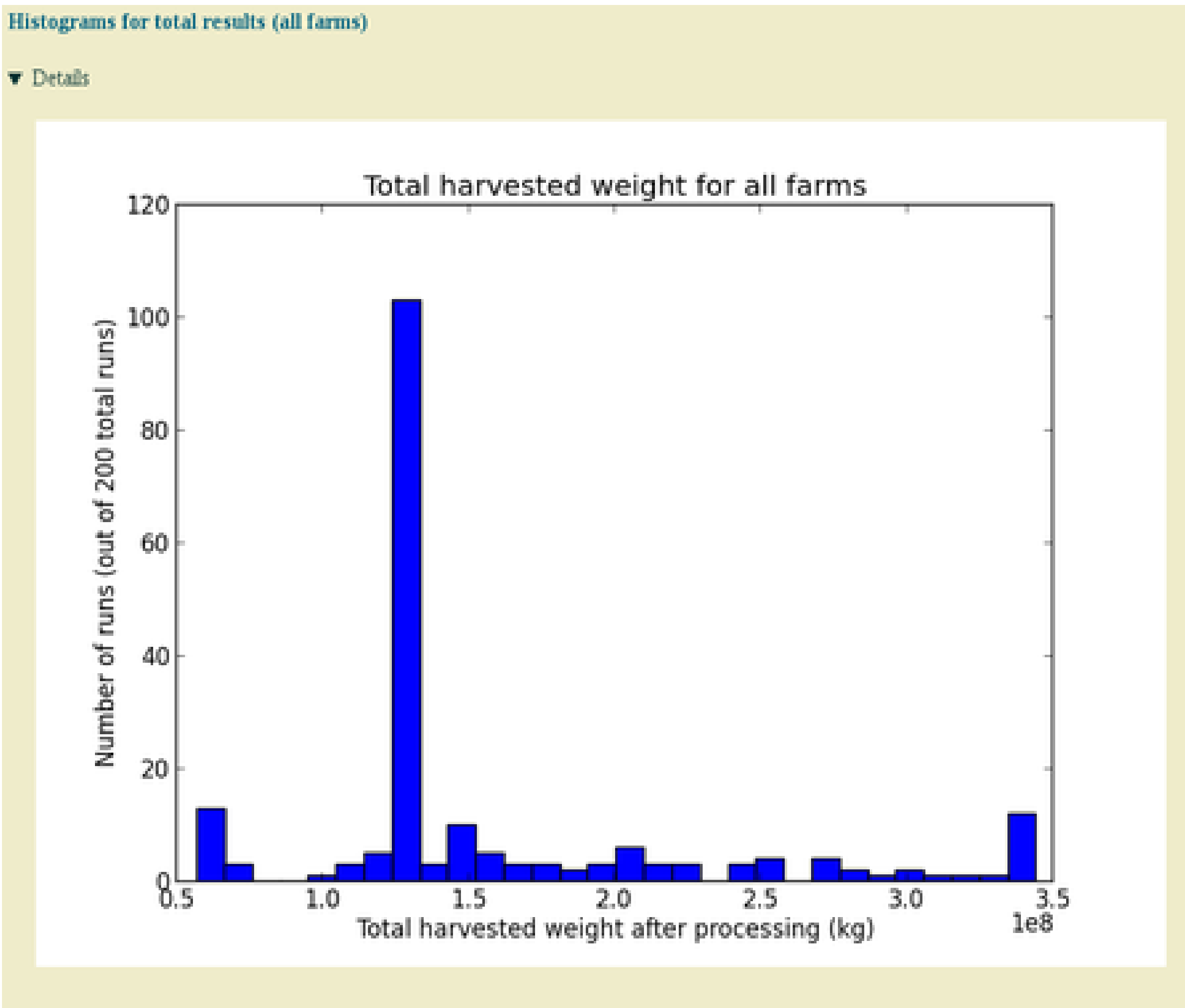


Fig. 5.44: Sample histogram in the uncertainty analysis section of HTML output

FAO. 2004. Fishstat Plus. Universal software for fishery statistical series. Capture production 1950 - 2004. FAO Fish. Aqua. Dept., Fish. Inf., Data, Stat. Dep.

Goldburg R., and R. Naylor. 2004. Future seascapes, fishing, and fish farming. Front. Ecol. 3:21–28.

Halwart, M., D. Soto, and J.R. Arthur, J.R. (eds.) 2007. Cage aquaculture – Regional reviews and global overview. FAO Fisheries Technical Paper. No. 498. Rome, FAO. 241 pp.

Naylor, R., and M. Burke. 2005. Aquaculture and Ocean Resources: Raising Tigers of the Sea. Ann. Rev. Envtl. Res. 30:185-218.

Soto, D., J. Aguilar-Manjarrez, and N. Hishamunda (eds). 2008. Building an ecosystem approach to aquaculture. FAO/Universitat de les Illes Balears Expert Workshop. 7–11 May 2007, Palma de Mallorca, Spain. FAO Fisheries and Aquaculture Proceedings. No. 14. Rome, FAO. 221p.

Stigebrandt, A., 1999. Turnover of energy and matter by fish—a general model with application to salmon. Fisken and Havet No. 5, Institute of Marine Research, Norway. 26 pp.

5.14 Fisheries

5.14.1 Summary

Wild capture fisheries provide a significant source of protein for human consumption and directly employ nearly 40 million fishers worldwide (FAO 2014). However, poor harvesting practices and habitat loss and degradation can reduce the ability of ecosystems to support healthy, productive fisheries. The InVEST Fisheries Production model produces estimates of harvest volume and economic value of single-species fisheries. The model is an age- or stage-structured population model, and is presented as a generic model that can be adapted to most species and geographies. Inputs to the model include parameters for life history characteristics (e.g., age at maturity, recruitment, migration and natural mortality rates), behavior of the fishery (e.g., fishing pressure), habitat dependencies (e.g., importance and availability of nursery habitat), and, optionally, economic valuation (e.g., price per unit biomass). The model outputs the volume and economic value of harvest within the area(s) designated by the user. It is best to compare outputs from multiple runs of the model, where each run represents different scenarios of habitat extent, environmental conditions and/or fishing pressure. A library of four sample models is provided, which the user can adapt to their own species or region, or the user can choose to build a model from scratch.

5.14.2 Introduction

Marine and aquatic ecosystems provide habitat for fish and shellfish, which in turn provide food and livelihoods for millions of people worldwide (FAO 2014). The ability of ecosystems to support fisheries depends on having intact habitat for fish, and on maintaining harvests at sustainable levels. A consideration of how changes in habitat or harvesting practices will impact the production of wild fish is thus important when weighing decisions which impact marine or aquatic ecosystems.

The status and ecology of fish stocks is often assessed by compiling multiple types of data into a single model that gives estimates of production under different scenarios. Unfortunately, such complex stock assessments are often not possible due to a lack of data and/or resources. In addition, traditional stock assessments generally do not take into account habitat dependencies or spatial dynamics, both of which are essential for understanding how local or regional fisheries production might respond under different scenarios. Therefore, a tool is needed that is flexible enough so that it can be adapted to different species, localities, and qualities of data, and which can be used to assess the potential consequences of decisions on the production of wild capture fisheries.

5.14.3 The Model

The InVEST model of ecosystem services from fisheries is an age- or stage-structured, deterministic, population dynamics model for an individual species. The model uses life-history information and survival parameters provided by the user to estimate the volume of harvest. The model can then be used to explore how the amount of harvest (and, optionally, value) responds to changes in amount of habitat (e.g., seagrass, mangrove, coral reef), environmental conditions (e.g., temperature, salinity), and/or fishing pressure. It is best to compare outputs from multiple runs of the model, where each run represents different scenarios of habitat extent, environmental conditions and/or fishing pressure. Fish population dynamics are notoriously variable and difficult to predict. This model is not intended to give a precise prediction of harvest amounts, but rather to be used as a tool to explore the consequences of different decisions which could impact fisheries production.

Parameter sets for four sample models are provided, representing the following species and geographies: (1) Caribbean spiny lobster (*Panulirus argus*) in Belize; (2) Dungeness crab (*Metacarcinus magister*) in Hood Canal, Washington; (3) blue crab (*Callinectes sapidus*) in Galveston Bay, Texas; and (4) white shrimp (*Litopenaeus setiferus*) in Galveston Bay, Texas. We chose these combinations of species and geographies because they were of interest to our partners in different NatCap application sites. The existing models, and others that will be added as they are developed, capture a range of life history types and exploitation patterns such that users can choose an existing model and modify it for their own region and species (e.g., modify the Galveston Bay white shrimp model for brown shrimp in the South Atlantic). Alternatively, the model is formulated such that a user with more advanced knowledge of fisheries science and modeling techniques can start from scratch and parameterize the generic model to suit any species (or guild) of interest.

How it Works

The underlying mechanics of the model are an age-structured or stage-structure population dynamics model. The model interface guides users through a series of decisions about the model's structure:

- Are populations structured by *age* or by *stage*?
- Should males and females be modeled *separately (sex-specific) or together*?
- Are there multiple *subregions*?
- Is there *migration* between subregions?
- How is *recruitment* (i.e. the production of offspring) determined?
- For *how many time steps* should the model run?

The user supplies the necessary parameters, which describe the survival rates, maturation schedule, recruitment function, migration patterns, and vulnerability to harvest. The model then runs for a user-specified number of time steps with the intention of the population reaching a state of equilibrium. Primary model outputs are estimates of harvest and value (optional) for the population at the final time step of the run. *Valuation* is optional and reflects revenue earned from sale of processed catch.

After generating a baseline model run, the user can then alter aspects of the model to compare fisheries production under different scenarios. Scenarios feed into the model by altering survival rates at certain life stages or in certain locations, for instance in response to changes in habitat extent, environmental variables, and/or fishing. An optional *Habitat Scenario Tool* is provided to assist the user in generating new survival parameters based on changes in habitat area, such as a decrease in the amount of eelgrass habitat or an increase in coral habitat.

Users have many options that can be chosen to customize the model to their particular species or question. We provide guidance for how to customize the model, as well as pointers to examples of model calibration and validation, both of which are done outside of the modeling framework.

Age-Structured Populations

Age-structured populations are modeled as:

$$N_{a,s,x,t} = \begin{Bmatrix} Rec_{s,x,t} & if & a=0 \\ \left(N_{a-1,s,x,t-1}Mig^x_{a-1,s,x} + \sum_{x'\neq x} N_{a-1,s,x',t-1}Mig^x_{a-1,s,x'}\right) S_{a-1,s,x} & if & 1 \leq a < A \\ \left(N_{A-1,s,x,t-1}Mig^x_{A-1,s,x} + \sum_{x'\neq x} N_{A-1,s,x',t-1}Mig^x_{A-1,s,x'}\right) S_{A-1,s,x} & if & a = A \\ +\left(N_{A,s,x,t-1}Mig^x_{A,s,x} + \sum_{x'\neq x} N_{A,s,x',t-1}Mig^x_{A,s,x'}\right) S_{A,s,x} & & \end{Bmatrix}$$

Where $N_{a,s,x,t}$ is the number of individuals of age *a* (*A* = maximum age) of sex *s* in area *x* at the start of time step *t*; and $S_{a-1,s,x}$ is survival from natural and fishing mortality from age *a*-1 to *a* for each sex and area; $Rec_{s,x,t}$ is recruitment of new individuals/number of offspring; $Mig^x_{a,s,x'}$ is the proportion of individuals of age *a* sex *s* that migrate from area *x'* to area *x* (or the proportion that remain in the area if *x'* = *x*).

Survival from natural and fishing mortality is defined as:

$$S_{a,s,x} = surv_{a,s,x}\left(1 - Ex_x * V_{a,s}\right)$$

Where $surv_{a,s,x}$ is survival from natural fishing mortality from age *a* to *a* + 1. for each sex and area; Ex_x is exploitation, which is the proportion of the population vulnerable to harvest that is actually harvested; and $V_{a,s}$ is vulnerability to harvest by age and sex. *Harvest* is assumed to occur at the beginning of the year, prior to mortality from natural causes.

Stage-Structured Populations

The stage-structured population model is a slightly more complicated version of the age-structured model, because we must account for variable length of the stages (e.g., stage 1 may last longer than 1 year, while stage 2 lasts less than one year). The model is as follows:

$$N_{a,s,x,t} = \begin{Bmatrix} \left(N_{a,s,x,t-1}Mig^x_{a,s,x} + \sum_{x'\neq x} N_{a,s,x',t-1}Mig^x_{a,s,x'}\right) P_{a,s,x} \quad + \quad Rec_{s,x,t} & if & a=0 \\ \left(N_{a-1,s,x,t-1}Mig^x_{a-1,s,x} + \sum_{x'\neq x} N_{a-1,s,x',t-1}Mig^x_{a-1,s,x'}\right) G_{a-1,s,x} & if & 1 \leq a \\ +\left(N_{a,s,x,t-1}Mig^x_{a,s,x} + \sum_{x'\neq x} N_{a,s,x',t-1}Mig^x_{a,s,x'}\right) P_{a,s,x} & & \end{Bmatrix}$$

Where $N_{a,s,x,t}$, $Rec_{s,x,t}$, $Mig^x_{a,s,x}$ are the same as in the formulation for the age-structured model. Note that there is no A (maximum age) for stage-structured models because the models are designed to capture all of the life stages. $G_{a,s,x}$ is the probability of surviving from natural and fishing mortality and growing into the next stage for each sex and area; and $P_{a,s,x}$ is the probability of surviving from natural and fishing mortality and staying in the same stage for each sex and area.

$G_{a,s,x}$ is a function of survival from natural and fishing mortality ($S_{a,s,x}$ as defined above, except now expressed as per unit time (e.g. day) rather than per year) and stage duration, D_a:

$$G_{a,s,x} = \frac{{S_{a,s,x}}^{D_a}\left(1 - S_{a,s,x}\right)}{1 - {S_{a,s,x}}^{D_a}}$$

$P_{a,s,x}$ is also a function of survival from natural and fishing mortality and stage duration; it is defined as:

$$P_{a,s,x} = S_{a,s,x}\frac{1 - {S_{a,s,x}}^{D_a - 1}}{1 - {S_{a,s,x}}^{D_a}}$$

Recruitment

Recruitment (i.e. the production of offspring) can be modeled in one of four different ways. The *Beverton-Holt* and *Ricker* functions assume different forms of density-dependence in the recruitment function (i.e. the number of offspring per adult decreases as adult abundance or biomass increases). The *Fecundity* function assumes a constant reproductive rate for adults (i.e. no density dependence). The *Fixed Recruitment* function assumes that recruitment is constant and not dependent on the number of adults. The parameterization of each function is shown below. See the *Guidance* section for advice on choosing and parameterizing the recruitment function.

Beverton-Holt

$$Rec_{s,x,t} = \frac{LarvalDispersal_x}{SexSpecific} \cdot \frac{(\alpha \cdot Sp_t)}{(\beta + Sp_t)}$$

Ricker

$$Rec_{s,x,t} = \frac{LarvalDispersal_x}{SexSpecific} \cdot \left(\alpha \cdot Sp_t \cdot e^{-\beta \cdot Sp_t}\right)$$

Fecundity

$$Rec_{s,x,t} = \frac{LarvalDispersal_x}{SexSpecific} \cdot \left(\sum_{a,s,x} N_{a,s,x,t-1} Maturity_{a,s} Fecundity_{a,s}\right)$$

Fixed Recruitment

$$Rec_{s,x} = \frac{LarvalDispersal_x}{SexSpecific} \cdot Recruitment$$

If the model is sex-specific, $SexSpecific = 2$, or if the sexes are aggregated, $SexSpecific = 1$. In sex-specific models, recruits are split evenly between males and females. $LarvalDispersal_x$ gives the proportion of larvae that settle in area x for models with subregions. Spawners (Sp_t) can be expressed as either number or biomass. The number of spawners is the product of the number of individuals in each age (or stage) class for the entire study region and the proportion that are mature by age (or stage):

$$Sp_t = \sum_{a,s,x} N_{a,s,x,t-1} Maturity_{a,s}$$

The biomass of spawners is the product of number of individuals in each age (or stage) class for the entire study region, the proportion that are mature at each age (or stage) AND their weight at a given age (or stage):

$$Sp_t = \sum_{a,s,x} N_{a,s,x,t-1} Maturity_{a,s} W_{a,s}$$

Where $W_{a,s}$ is weight or biomass by age and sex.

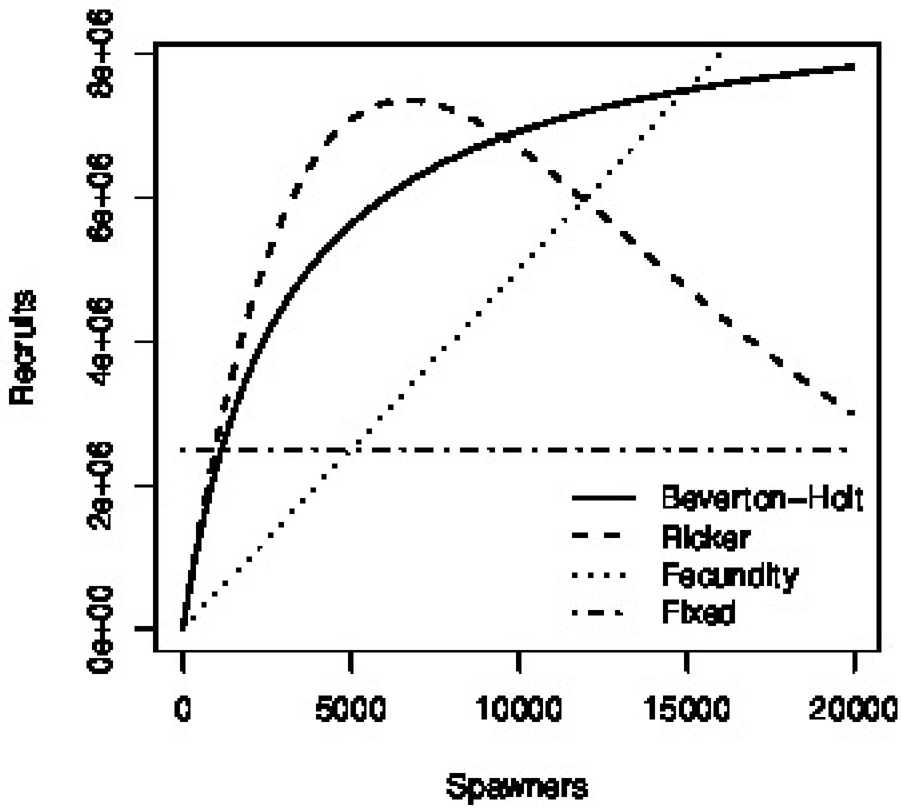


Fig. 5.45: Illustration of the four recruitment functions available within the InVEST Fisheries model.

Migration

If multiple subregions are specified, the user can choose to include migration from one subregion to another. Migration may occur at one or multiple ages/stages, and for each age/stage at which migration occurs the user specifies which proportion of fish from each subregion move to each other subregion. These proportions may depend on habitat quality, habitat quantity, known oceanographic dispersal patterns, etc. Within subregion migrations (for instance, shifts from one habitat type to another) should not be specified in this way. See Migration section in the Guidance section for more information.

Harvest

Harvest ($H_{x,t}$) from each subregion in the final (equilbrated) time step is calculated based on the user-defined exploitation rate(s) and vulnerability. Harvest can be output by numbers or by weight. Choosing the appropriate output metric depends on how catch is normally processed and sold.

Numbers (e.g. Dungeness crab: $H_{t,x} = \sum_{a,s,x} N_{a,s,x,t} Ex_x V_{a,s}$

By Weight (e.g. Spiny lobster, White shrimp): $H_{t,x} = \sum_{a,s,x} N_{a,s,x,t} Ex_x V_{a,s} W_{a,s}$

Where Ex_x is exploitation, which is the proportion of the population vulnerable to harvest that is actually harvested, $V_{a,s}$ is vulnerability to harvest for age a and sex s, and $w_{a,s}$ is weight for age and sex.

Valuation

Valuation, $V_{x,t}$ is optional and reflects the earnings from the sale of harvest. It is intended to give a rough idea of the current market value for an equilibrated population based on user-defined price parameters. It is simply:

$$V_{t,x} = H_{t,x} * FractionProcessed * Price$$

Where $Price$ is the value in price per units (where units match those given by $H_{x,t}$), and $FractionProcessed$ is the proportion of each harvest unit that remains to be sold after processing.

Initial Conditions

The user supplies the initial number of recruits for both age- and stage-structured models. To initialize the **age-structured** models the following is done (i.e., at $t = 0$):

$$N_{a,s,x,t=0} = \begin{Bmatrix} Rec_{s,x,t=0} & if & a = 0 \\ N_{a-1,s,x,t=0} S_{a,s,x} & if & 1 \leq a < A \\ \frac{N_{A-1,s,x,t=0} S_{A-1,s,x}}{(1-S_{A,s,x})} & if & a = A \end{Bmatrix}$$

For **stage-structured** models, we set the youngest stage as the initial recruitment, and then all other stages to 1 (as below). This is appropriate for the stage-structured models because each stage has a different duration, so we allow the model to redistribute the initial recruits (i.e., members of the youngest stage) over time instead of specifying them at the outset as in the age-structured model.

$$N_{a,s,x,t=0} = \begin{Bmatrix} Rec_{s,x,t=0} & if & a = 0 \\ 1 & if & 1 \leq a \end{Bmatrix}$$

Because the population model is run to equilibrium, the initial number of recruits will not affect the model results, but may affect the number of time steps required before the population reaches equilibrium.

5.14.4 Scenarios

The InVEST Fisheries model is best suited for comparing fisheries production under different scenarios. A scenario could be a change in the amount of juvenile habitat, a change in the harvest rate in a particular subregion, or a change in survival due to other causes such as climate change. Results from running scenarios can then be compared to baseline model runs to evaluate the consequences of such changes for fisheries production. To facilitate the analysis of scenarios, we provide a preprocessor tool for calculating how changes in habitat extent translate into changes in age/stage-specific survival.

Habitat Dependency

For ages/stages that depend on certain habitats (for instance, mangroves), a change in habitat coverage within a region can result in a change in the survival rate of ages/stages which depend on that habitat. The option to model this dependency is included as a *Habitat Scenario Tool* with InVEST, whereby new survival parameters are generated based on the baseline survival parameters and the amount of change in habitat. Users may choose to use the functional form provided in the tool, or use their own methods to calculate changes in survival. Currently, the tool is only suited for use with age-structured (not stage-structured) models.

Using the Habitat Scenario Tool, changes in the area of critical habitats are linked to changes in survival as follows:

$$S_{a,x} = surv_{a,x}\left(\frac{\sum_{d_{a,h}>0}\left(1+\frac{H_{h,x,SCEN}-H_{h,x,BL}}{H_{h,x,BL}}\right)^{d_{a,h}\gamma}}{n_a}\right)^{T_a}$$

Where $surv_{a,x}$ is baseline survival from natural mortality from age a-1 to a in subregion x: $surv_0 = 1$, $surv_a = e^{-M}$ if $a > 0$, and M_a is the natural mortality rate from a - 1 to a. T_a indicates if a transition to a new habitat happens from a - 1 to a, which is used so that changes in habitat coverage only affect survival during the transition to that habitat, but not once settled in the habitat. $H_{h,x}$ is the amount of habitat h (e.g. coral, mangrove, seagrass) in the region in the baseline (BL; i.e. status quo) system or under the scenario being evaluated (SCEN). $d_{a,h}$ is the degree to which survival during the transition from a-1 to a depends upon availability of h, y is a shape parameter which describes the relationship between a change in habitat and a change in survival, and n_a is the number of non-zero habitat-dependency values for age a. If $n_a = 0$, $S_{a,x} = surv_{a,x}$. $S_{a,x}$ is restricted to a maximum of 1.

5.14.5 Limitations and Simplifications

The InVEST Fisheries model is best suited for exploring how different scenarios of habitat change, harvesting, or changing environmental conditions may result in changes to fisheries production. It is not intended to be a stock assessment tool, nor should the output be interpreted as predictions of future catches. Fish populations are notoriously variable, both from year to year as well as over long time scales. In the InVEST Fisheries model, as with any model, the quality of the output will be determined by the quality of the parameters supplied. Key assumption of the model include:

- Fishing is assumed to take place at the start of the year, before natural mortality.
- After recruitment, survival is not density-dependent (i.e. does not depend on population size).
- Harvest rates and selectivity are fixed through time, such that technological improvements to gear or changes in fishing practices are not modeled.
- Market operations are fixed, such that they do not vary in response to amount of harvest, shifts in market or consumer preference.

Key assumptions of the Habitat Scenario Tool include:

- Habitat dependencies are obligatory (i.e., habitat substitutability is not explicitly represented).

- The population responds to change in habitat quantity (e.g., areal extent of mangrove, seagrass, or coral reef), not quality of those habitats.
- A change in habitat area affects survival only during the first life stage which depends on that habitat.
- Habitat availability is a limiting factor for survival. This means that an increase or decrease in the amount of habitat will always result in an increase or decrease in survival for the first life stage dependent on that habitat.
- The effect of a change in habitat on survival does not depend on the population density. In other words, a 50% reduction in juvenile habitat will have the same effect on survival rates regardless of the number of juveniles.

5.14.6 Model Details and Guidance

Customing the Model

Four sample models are included with the InVEST Fisheries model as parameter sets the user can input to InVEST. However, it is expected that users will customize the model to suit their own species or region as needed. The following sections provide guidance on how to customize the model, and give examples from the four sample models. For more information on the parameterization of the Dungeness crab model, as well as an application of the model, see Toft et al. 2013. For the Spiny lobster model used in the Belize case study, see Arkema et al. *in review* and Toft et al. *in prep* (available upon request).

Most of the parameters required for customizing the model may be found in scientific literature or reports, or based on local knowledge (e.g., maturation age or migration patterns). However, some parameters will likely need to be estimated from data (e.g., recruitment parameters). For instance, the Spiny lobster model was parameterized by fitting to time-series of catch and catch-per-unit effort (CPUE) from Belize (see Arkema et al. *in review*). This must be done outside InVEST and requires a user to be familiar with fitting models to data to estimate parameters. Some fisheries science expertise is also necessary.

As additional models are developed for particular applications, parameter sets and relevant files will be made publicly available and highlighted on the NatCap forum. User-developed models may be shared in the same way, with the vision of a growing library of InVEST Fisheries models from around the globe.

Guidance

Age or Stage Structured

An age-structured model is simply a stage-structured model where all stages are the same length (typically one year). If multiple important life-history transitions happen within a year which should be captured in the model (e.g., multiple transitions from one habitat to another, or multiple migration events between regions), then a stage-structured model may be most appropriate.

All of the sample models are age-structured models, aside from shrimp, which is stage-structured. The stage-structured model accounts for variable lengths of the stages (e.g., stage 1 may last longer than a year, while stage 2 may last less than a year). Stage duration, D_a, must be specified for each stage, but is assumed to be constant (and typically 1 year) for age-structured models.

Time Step Units

For age-structured models, the time step is assumed to be one year, and parameters are therefore based on annual rates and the model progresses in one-year increments. For stage-structured models, the user determines the time step. For instance, in the white shrimp model time steps are interpreted as days because 'Duration' values in the population_params.csv are number of days. The time step unit (days, months, years) will be the same as used for the "number of time steps for model run," which is specified by the user.

Number of Time Steps for Model Run

The number of time steps should be sufficiently large for the population to reach equilibrium. For age-structured models, a reasonable starting place is between 100-300 time steps, but will depend on the population parameters. For stage-structured models, more time steps may be needed. It is recommended that the user start with an intermediate number of time steps and check the model output to determine whether more time steps are needed to reach equilibrium.

Number of Age or Stage Classes

The model should capture the major points through adulthood– larval, juvenile/rearing, spawning, harvest. This is fairly straightforward for stage-structured models, as stages will span larval to adult stages, with some in between. For white shrimp, for instance, the intermediate stages are post-larval, marsh and bay. For age-structured models, the maximum age should be set to be old enough for the species to have reached full maturity and to be subjected to maximum harvest. The oldest age class will be a 'plus' class meaning that it includes that age and all older ages. There is no maximum age for stage-structured models because the models are designed to capture all of the life stages.

Spiny Lobster (8 age classes): 0 (larval), 1, 2, 3, 4, 5, 6, 7+

Dungeness Crab (5 age classes): 0 (larval), 1, 2, 3, 4+

Blue Crab (4 age classes): 0 (larval), 1, 2, 3+

White Shrimp (5 stage classes): eggs/larvae, post-larval, marsh, bay, adult (based on Baker et al. 2008)

Sex-Specific or Not

A sex-specific model can be used if the biology (e.g., migration, size at age) or harvest practices differ substantially by sex. Different parameters can be given to each sex. Of the sample models, Dungeness crab is the only example of a sex-specific model. Males and females are separate in the Dungeness crab model because regulations prohibit harvest of female crabs. The population model could have been combined for both sexes, but we deemed it easier to keep them separate to reflect the harvest practices.

Areas(s) of Interest

The model can encompass one area—that is, be completely spatially aggregated (i.e. the population is considered homogenous throughout the study area)—or the area can be divided into subregions. In the sample models, we have made our decisions about how to include space in each model based on the policy questions and data availability for parameterizing the model.

Spiny Lobster: Project partners separated Belizean coastal and marine waters into 9 planning regions of different sizes, which we use for the lobster model (Clarke et al. 2013).

Dungeness Crab: Six boxes of irregular shape/size to match output from an ecosystem model (Toft et al. 2013).

Blue Crab and White Shrimp: A single bay-wide region (Guannel et al. 2014).

Larval Dispersal

For models with subregions (e.g., Spiny Lobster, Dungeness Crab), we assume that adults from each subregion contribute to a common larval pool. Larvae are then distributed across subregions. The proportion of larvae that go to each subregion is user-defined, in the main parameters csv file. In the spiny lobster default model, larvae are dispersed to the subregions according to the distribution of suitable habitat (e.g. mangroves and seagrasses) among the subregions

(Arkema et al. *in review*). In the Dungeness crab default model, larvae are dispersed proportional to the surface area of each subregion (Toft et al. 2013). The models represent closed populations, meaning we do not allow for any larval recruitment from outside of the study area. However, if recruitment is modeled using the *Fixed Recruitment* function, this could implicitly represent an external source of larvae.

Migration

If there are multiple subregions in the model, the user defines the degree of migration between subregions and at what ages/stages this migration occurs. In deciding how to include migration in a model, the user should answer questions such as: Does a portion of each age-class emigrate each year? Or does emigration only occur for specific classes as they migrate between habitats? What portion of each class emigrates (e.g., 10% or 50%)? Where do they go (e.g., distribute equally to all other subregions regardless of distance from subregion of origin, or distribute based on a distance decay from subregion of origin)?

To specify migration, the user includes a separate matrix for each age (or stage) when migration occurs (e.g., in the lobster model, lobster migrate between ages 2 and 3 only, so only 1 migration matrix is included). These matrices, stored within a single folder, are selected under "migration matrix CSV folder" in the model interface. Note that movements within subregions (for instance, ontogenetic shifts between different habitat types) may be implicitly included in the model by altering age-specific survival rates to reflect availability of recipient habitat (see habitat dependency section). Within subregion movements do not require a migration matrix.

Spiny lobster is the only sample model that includes migration, which occurs as lobster move from mangroves and seagrasses to corals between ages 2 and 3. The proportion of age 2s that migrate from one subregion to another is determined by a distance decay function weighted by the amount of coral habitat in each subregion. For example, if there are 2 subregions and one is replete with coral, more of the age 2 lobster will migrate to that subregion than the other (for details see Arkema et al. *in review*).

Survival from Natural Mortality

Each year, a proportion of each age-class or stage succumbs to natural mortality due to a variety of causes, including predation, disease, or competition. Survival from natural mortality is the proportion of individuals that continues on to the next age/stage. Often, survival from natural mortality is calculated from instantaneous natural mortality rates (M_a), which are frequently available from peer-reviewed literature and/or stock assessments: $S_a = e^{-M_a t}$, where t is the length of the time step over which survival is calculated (typically 1 year for *Age-Structured Populations*).

Within the model, natural mortality may vary by age/stage, sex, and subregion, but it may not vary by time step.

Spiny Lobster: survival from natural mortality is the same across all ages (0.698), as calculated from a natural mortality rate of $M = 0.36y^{-1}$ (Puga et al. 2005).

Dungeness Crab: we use 4 survival parameters, which were the same for males and females (see Higgins et al. 1997, Toft et al. 2013, and references therein). The survival of eggs to age 1 crab involves survival through two phases of Dungeness crab development—egg, and megalopae—for which we each had estimates of survival (5.41x10-6 and 0.29, respectively). We multiplied these together to generate the survival term from eggs through megalopae to age 1. Survival was the same for ages 2 and 3 of both sexes, and age 4+ females (0.725); age 4+ males are harvested and the surviving males have been shown to have a lower survival than other adult Dungeness crab (0.526).

Survival from Fishing Mortality

Mortality from fishing depends on the exploitation fraction and the age- or stage-specific vulnerability to harvest (see below).

Exploitation Fraction

This is the proportion of the population vulnerable to harvest that is actually harvested. This may vary by subregion.

Vulnerability to Harvest

Not all ages or stages are equally likely to be harvested. Vulnerability to harvest (also called selectivity) may depend on size, life-stage specific behavior (for instance spawning aggregations), habitat use, or regulations, and may change depending on the gear and fishing strategies employed. A value of 1.0 indicates that the age or stage is fully vulnerable to harvest, whereas values less than one indicate the vulnerability relative to the fully-vulnerable age or stage. For instance, if all individuals age 4+ are fully vulnerable, whereas age 3 individuals are only half as likely to be caught given the same fishing pressure, age-3 would have a vulnerability of 0.5. The most vulnerable age/stage should have a value of 1.0. Vulnerability is assumed to be the same across subregions.

Different functional forms may be used to describe vulnerability. These are examples intended to help the user construct the population parameters csv file, but other functional forms are possible (for instance, a dome-shaped function would imply the highest vulnerability for medium-aged individuals).

Binary: each age or stage is either not vulnerable or fully vulnerable (0 or 1).

Logistic function: assumes that vulnerability increases with age/stage, where a_{50} is the age at which individuals have a 50% vulnerability to harvest, and δ determines the slope of the logistic function.

Spiny Lobster: We model vulnerability-at-age by using the logistic function above, with a_{50} set to 2.5 years and δ set to 10. A δ of 10 gives the shape of the logistic function a nearly knife-edge selectivity, meaning that very few lobster younger than 2.5 years are vulnerable to fishing, whereas almost all lobster older than 2.5 years are vulnerable to fishing. This cutoff was chosen as this is the age when lobster reach the minimum legal size for harvest of 75mm. A smaller delta would soften the knife-edge selectivity, resulting in higher vulnerability (and harvest) of younger lobster. Exploitation (Ex_x) for this model is set to 31% based on historical harvest rates.

Dungeness Crab: Vulnerability and exploitation are set more simply in this model. Only age 4 males are assumed to be vulnerable to harvest (V = 1 for age 4 males, and V = 0 for all other ages and females). Ex_x is set to 0.47, meaning 47% of age-4 males are harvested in each region. This was estimated by adjusting an average harvest rate for California, Oregon and Washington to include only tribal and recreational catch since commercial harvesting does not occur in Hood Canal, WA (details in Toft et al. 2013)

Recruitment

Beverton-Holt: The Beverton-Holt model represents a situation where the total number of recruits increases with spawners abundance up to an asymptote. This recruitment function also has two parameters: alpha and beta. For Beverton-Holt, alpha represents the maximum number of recruits produced (i.e. the asymptote), whereas beta represents the number of spawners needed to produce recruitment equal to half the maximum (alpha/2). In this form, alpha/beta represents the recruits per spawner at low spawner levels.

Ricker: The Ricker model represents a situation where the total number of recruits increases up to intermediate spawner levels and then decreases at very high spawner levels. This function has two parameters: alpha and beta. In the Ricker model, alpha gives the maximum recruits per spawner at low spawner levels (i.e., the initial slope of the stock-recruit curve), while beta is the rate of decline in recruits as there are more spawners, or the degree to which the curve bends downwards as spawner abundance increases.

For both Ricker and Beverton-Holt, spawners may be measured in numbers of individuals or in biomass, and the parameters should be specified appropriately.

Fecundity: For the fecundity-based recruitment function, only age- or stage-specific fecundity values are needed, representing the number of offspring per mature individual. Caution is urged when selecting this option as age-based models must be carefully parameterized in order to reach equilibrium. Most parameter sets will result in a continuously increasing or decreasing population. We do not recommend this option for stage-based models.

Fixed: In the fixed recruitment function, recruitment is time-invariant. A value for the fixed number of recruits must be given. Recruitment therefore does not depend on the abundance of mature individuals.

Note: Choosing which recruitment function to use will depend on data availability as well as ecological knowledge about the species and region. Density-dependent recruitment functions such as the Ricker and Beverton-Holt are most common in fisheries models, as they recognize that a population depends on finite resources and cannot grow infinitely large. A model with the Fecundity function must be parameterized carefully or it is not guaranteed to reach an equilibrium. The Fixed recruitment may be appropriate in cases where the region of interest is small relative to the range or distribution of the fished population, for instance, when recruits may drift into the region of interest from nearby spawning areas.

The Ricker function is used for the blue crab and Dungeness crab models. The lobster model uses the Beverton-Holt function. The white shrimp model assumes fixed recruitment. In all cases, stock-recruitment parameters were estimated by fitting the model to available data. For instance, the spiny lobster model was fit to three time-series of catch-per-unit-effort (CPUE) data, which allowed the estimation of alpha and beta. In the white shrimp model, recruitment was estimated by fitting the model to catch data.

If the user would like to create their own recruitment function for the Fisheries Model, an optional parameter has been created in the InVEST Fisheries python module (but not in the User Interface) to allow for this. See the Fisheries Model page of the InVEST API Reference for more information.

Initial Recruitment

Because the model is an equilibrium model, the value chosen for Initial Recruitment is not critical. It should be in a reasonable range in order to ensure the model reaches equilibrium without too many time steps.

Maturity at Age/Stage

Maturity at age or stage is used to determine the abundance of spawners if you choose Beverton-Holt or Ricker for the recruitment function. These parameters may be taken from other studies, or estimated from data using, for instance, the equation given below. If local data are not available, www.fishbase.org provides basic life history information for many species of fish.

Maturity-at-age, m_a, may be calculated using a maturity ogive governed by a logistic function:

$$m_a = \left(1 + exp\left(-\phi\left(L_a - L_{50}\right)\right)\right)^{-1}$$

where:

- ϕ determines the slope of the logistic function
- L_{50} is the length-at-50% maturity
- L_a is the length-at-age, defined according to a von Bertalanffy growth equation.

See the length/weight section below.

Weight at Age (Optional)

It is optional to include a length-weight relationship in the model. We have done so for the lobster model because harvest and sale of lobster is recorded by volume of meat, not number of lobsters, which means any model validation needed to be a comparison of weight, not just numbers of lobster. For Dungeness crab, however, we validated the model against numbers of crab landed and had no need to transform numbers of crab to volume of crab. It may also be useful to include weight when recruitment depends strongly on total biomass of spawners, rather than total numbers of spawners.

For Lobster, we use a von Bertalanffy growth equation and a length-weight relationship to transform numbers-at-age to weight-at-age, w_a:

$$w_a = e{L_a}^f$$

where:

- e, f are parameters of the von Bertalanffy growth equation
- L_a is length-at-age

L_a is defined as:

$$L_a = l_\infty \left(1 - exp\left(-\kappa \left(a - t_0\right)\right)\right)$$

where:

- I_∞ is the asymptotic maximum length
- κ is curvature parameter, which is proportional to rate at which I_∞ is reached
- t_0 is the age at which the fish has 0 length, therefore is non-negative, or zero.

Parameters for these equations may be estimated from data prior to running InVEST, or taken from www.fishbase.org for many fish species. Alternatively, estimates of weight-at-age may be taken directly from fish measurements without using a model.

Valuation (Optional)

Valuation is intended to reflect the earnings from the sale of harvest. *Unit Price* gives the price per unit of harvest (either weight or numbers) that fishers receive from buyers. This information should be obtainable in reports, from national statistics, or by surveying fishers and buyers. *Fraction Kept After Processing* gives the proportion of each unit of harvest that remains to be sold after processing, or if harvest is specified in numbers, gives the proportion of individuals which are sold.

Currently, the spiny lobster model is the only sample model with valuation. See Arkema et al. *in review* and Toft et al. *in prep (available upon request)* for a description of how valuation parameters were estimated.

Habitat Dependency and the Habitat Scenario Tool (Optional)

Habitat dependencies are not explicitly included within the InVEST Fisheries Production model. However, a *Habitat Scenario Tool* is included which can be used to generate updated survival parameters based on changes in habitat. In order to use the tool, the user must already have a baseline set of population parameters, particularly survival rates. The tool takes information on changes in habitat area (expressed as a percent change from the baseline habitat area), age/stage specific habitat dependencies (i.e. age-0 lobster depend on mangroves and seagrass), and a user-specified shape parameter describing the relative response rate, or how a change in habitat corresponds to a change in survival. The tool outputs a new population parameters file with updated survival rates. Note that this tool cannot be used to generate an initial set of survival parameters, but is only used to update baseline survival parameters based on habitat change scenarios.

In using this tool, the user should have information on which life stages depend on which habitat types.

The user needs to specify:

1. Habitat changes, represented as a percent change in the area of each habitat type by subregion (if applicable). Changes in habitat area can represent different scenarios of conservation, restoration, or development, for instance as output from the Habitat Risk Assessment model.
2. Age- or stage-specific habitat dependencies, ranging from 0 (no dependency) to 1 (fully dependency). If an age or stage depends on multiple habitats, each habitat-stage dependency value can range from 0 to 1. However, if habitats are interchangable with regards to species dependency (in other words, if a species can use either habitat type, or if an increase in one habitat can compensate for a decrease in the other), we recommend modeling them as a single habitat type in the *Habitat Scenario Tool*. Information on habitat dependencies can often be found in the scientific literature.
3. A gamma value. A gamma value of 1 means that a 50% increase in habitat area will correspond to a 50% increase in survival. A gamma value of 0.2 means that a 50% increase in habitat area will correspond to only a 10% increase in survival.

5.14.7 Data Needs

Many types of data may and should be used to estimate inputs for the model parameters. For instance, data about a species' length, weight, maturity, or fecundity at a given age/stage are important for specifying how the population reproduces. Historical data on prices can be used to estimate the value of harvests. Survival rates may be estimated from data or taken from literature values. Because the types of data available for each fishery may vary drastically, the model is designed to allow the user full flexibility in how these inputs are estimated. In cases where parameters are highly uncertain, we recommend the user to run the model multiple times with a range of parameter values to determine how sensitive the model's results are to uncertainty in parameters. For more details on the definitions of the input data, please see the *How it Works* and *Guidance* sections.

5.14.8 Running the Model

Core Model

Upon opening the Fisheries program, the user is presented with an interface containing a set of parameters through which to submit inputs. Information about each parameter is provided below. Once the user has entered all necessary inputs, the user can start the model run by pressing 'Run'. If any errors occur, InVEST will stop the model run and provide feedback to the user about what caused the error through a message screen.

General Parameters

1. **Workspace (required)**. The selected folder is used as the workspace where all intermediate and output files will be written. If the selected folder does not exist, it will be created. If datasets already exist in the selected folder, they will be overwritten.

   Naming Conventions: Any alphanumeric string of characters. Best to avoid whitespace characters.

   Example Filepath: \InVEST\Fisheries\

2. **Area(s) of Interest (optional)**. The provided shapefile is used to display outputs within the subregion(s) of interest. The layer should contain one feature for every subregion of interest, each feature of which should have a 'Name' attribute (case-sensitive) matching a corresponding subregion in the Population Parameters CSV File. The 'Name' attribute value can be numeric or alphabetic, but must be unique within the given file.

   Filetype: Polygon Shapefile (SHP)

   Example Filepath: \InVEST\Fisheries\Input\lobster_subregions.shp

   Requirement: must have a 'Name' attribute in the shapefile's attribute table.

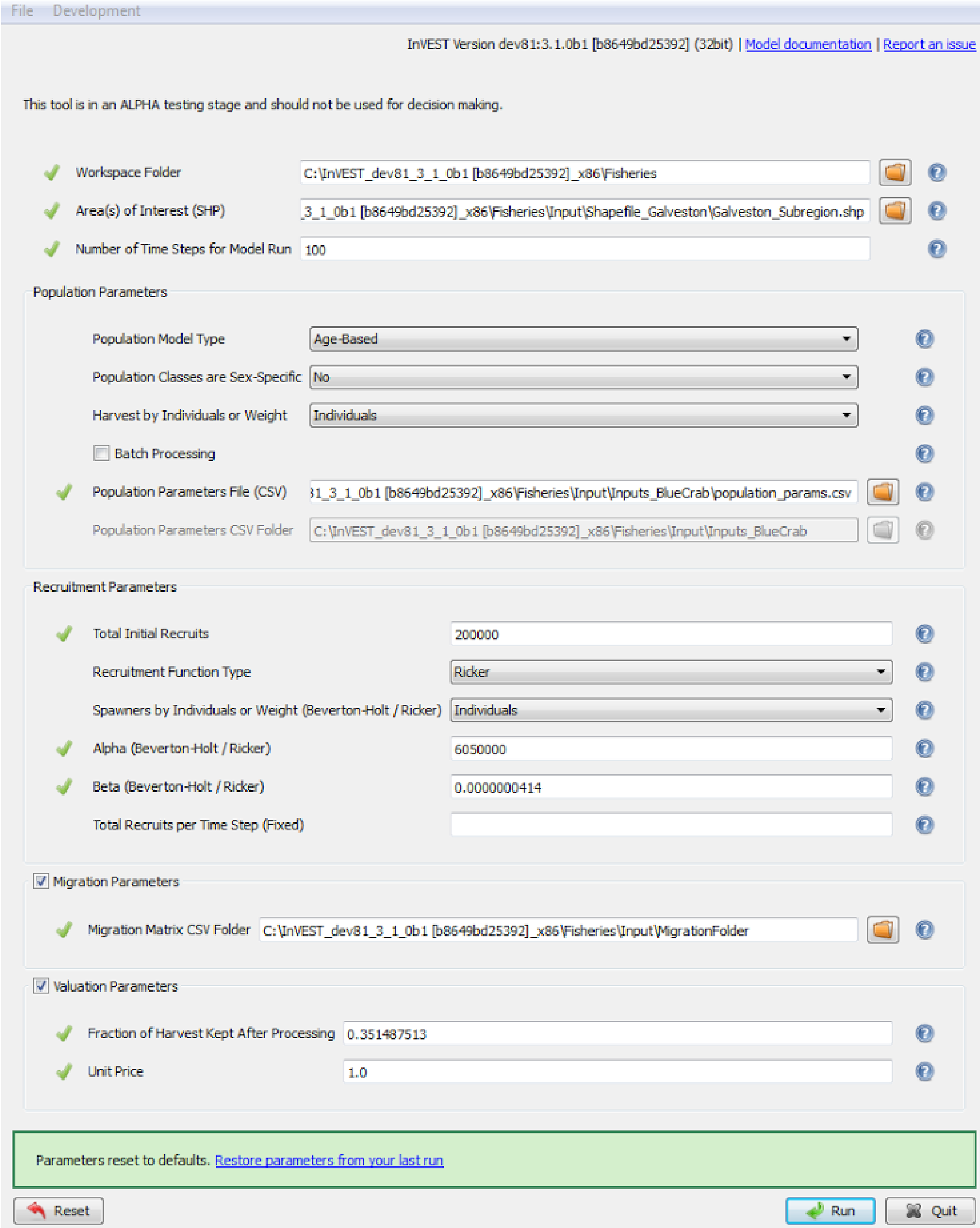


Fig. 5.46: Example User Interface for Core Model

3. **Number of Time Steps for Model Run (required)**. The number of time steps the simulation shall execute before completion. Must be a positive integer. The time step can use any unit of time relevant to the population. Consult *Time Step Units* for advice on selecting time step duration.

Population Parameters

4. **Population Model Type (required)**. Specifies whether the classes provided in the Population Parameters CSV file represent ages or stages. Age-based models (e.g. Spiny Lobster, Dungeness Crab) are separated by uniform, fixed-length time steps (usually representing a year). Stage-based models (e.g. White Shrimp) allow stages to have non-uniform durations based on the assumed resolution of the provided time step. If the stage-based model is selected, the Population Parameters CSV File must include a 'Duration' vector alongside the survival matrix that contains the number of time steps that each stage lasts.
5. **Population Classes are Sex-Specific (required)**. Specifies whether or not the population classes provided in the Population Parameters File are distinguished by sex.
6. **Harvest by Individuals or Weight (required)**. Specifies whether the harvest output values are calculated in terms of number of individuals or in terms of biomass (weight). If 'Weight' is selected, the Population Parameters CSV File must include a 'Weight' vector alongside the survival matrix that contains the weight of each age/stage, as well as sex if the model is sex-specific.
7. **Batch Processing**. Specifies whether the program will perform a single model run or a batch (set) of model runs. For single model runs, users submit a filepath pointing to a single Population Parameters CSV file. For batch model runs, users submit a folder path pointing to a set of Population Parameters CSV files. The name of each CSV file will serve as the prefix of the outputs created by the model run.
8. **Population Parameters CSV File**. The provided CSV file should contain all necessary parameters for population groups based on age/stage, sex, and subregion - excluding possible migration parameters.

   Naming Conventions: Any alphanumeric string of characters. Best to avoid whitespace characters.

   Filetype: Comma Separated Values (CSV)

   Example Filepath: \InVEST\Fisheries\Inputs_Lobster\population_params.csv

   - **Classes (required)**- The leftmost column should contain the age/stage names of the given species listed in chronological order. Each name can be an alphanumeric string. If the population classes are sex-specific, all age/stage names of one sex must be listed first, followed below by the age/stage names of the other sex.
   - **Subregions (required)**- The top-most row should contain the subregion names considered by the model. Each name can be an alphanumeric string. If the AOI shapefile is to be provided, the subregion entries must each match a corresponding 'Name' attribute value in a feature of the AOI shapefile. An entry must be provided even if the model is considering only one subregion.
   - **Survival Rates from Natural Mortality Matrix (required)**- Each unique pair of age/stage and subregion should contain a survival rate from natural mortality, expressed as a decimal fraction.

   Subregion-specific Attributes: Rows placed directly below the survival matrix with at least one empty row placed in-between as a buffer.

   - **ExploitationFraction (required)**- A row starting in the first column with the label 'ExploitationFraction'. The exploitation fraction is the proportion of the vulnerable population in each subregion that is harvested (0=0% harvested, 1=100% harvested). Each subregion is treated independently (i.e. up to 100% of the vulnerable population in each subregion may be harvested).
   - **LarvalDispersal**- A row starting in the first column labeled 'LarvalDispersal'. The larval dispersal is the proportion of the cumulative larvae pool that disperses into each subregion. Each subregion column should have a decimal to represent this. Dispersal across all subregions should add up to 1. If larval dispersal isn't provided, larvae will be dispersed equally across all subregions.

Class-specific Attributes: Columns placed directly to the right of the survival matrix with at least one empty column placed in-between as a buffer.

- **VulnFishing (required)**- A column labeled 'VulnFishing', which is the relative vulnerability to harvest for each class. A decimal value for each class listed in this column is required. The most vulnerable age(s)/stage(s) should have a value of 1.0, indicating full vulnerability.
- **Maturity**- A column labeled 'Maturity'. This column is only required if the recruitment function being used is Ricker, Beverton-Holt, or Fecundity. It represents the fraction of that age or stage which is mature and contributes to the spawning stock. A decimal value for each age/stage is required if maturity is included. For classes which do not reproduce, this should be 0.
- **Duration**. A column labeled 'Duration'. This column is required for stage-based models. It represents the number of time steps for which an average individual will be in that stage before moving to the next one.
- **Weight**- A column which is required if 'Spawners by Weight' or 'Harvest by Weight' is selected. This is the average biomass of an individual of the population at each age/stage expressed in model-agnostic units, and is required for each of the ages/stages listed in the classes column.
- **Fecundity**- A column in the headers row which is required if the recruitment function being used is Fecundity. It represents the number of recruits per mature individual.

Example Sex-Aggregated Population Parameters CSV File

Class	Subregion_1	Subregion_2	...	Subregion_N		**VulnFishing**	**Maturity**	**Duration**	**Weight**	**Fecundity**
Class_1	<float>	<float>	...	<float>		<float>	<float>	<int>	<float>	<float>
Class_2	<float>	<float>	...	<float>		<float>	<float>	<int>	<float>	<float>
...	...	...	...	...		...	...	...	...	...
Class_N	<float>	<float>	...	<float>		<float>	<float>	<int>	<float>	<float>
ExploitationFraction	<flaot>	<float>	...	<float>						
LarvalDispersal	<float>	<float>	...	<float>						

Example Sex-Specific Population Parameters CSV File

Class	Subregion_1	Subregion_2	...	Subregion_N		**VulnFishing**	**Maturity**	**Duration**	**Weight**	**Fecundity**
Class_1 (Female)	<float>	<float>	...	<float>		<float>	<float>	<int>	<float>	<float>
Class_2	<float>	<float>	...	<float>		<float>	<float>	<int>	<float>	<float>
...	...	...	...	...		...	...	...	...	...
Class_N	<float>	<float>	...	<float>		<float>	<float>	<int>	<float>	<float>
Class_1 (Male)	<float>	<float>	...	<float>		<float>	<float>	<int>	<float>	<float>
Class_2	<float>	<float>	...	<float>		<float>	<float>	<int>	<float>	<float>
...	...	...	...	...		...	...	...	...	...
Class_N	<float>	<float>	...	<float>		<float>	<float>	<int>	<float>	<float>
ExploitationFraction	<float>	<float>	...	<float>						
LarvalDispersal	<float>	<float>	...	<float>						

9. **Population Parameters CSV Folder**. The provided CSV folder should contain a set of Population Parameters CSV files with all necessary attributes for population classes based on age/stage, sex, and subregion – excluding possible migration information. The name of each file will serve as the prefix of the outputs created by the model run.

Recruitment Parameters

10. **Initial Number of Recruits (required)**. The initial number of recruits in the population model at time equal to zero. If the model contains multiple subregions of interest or is distinguished by sex, this value will be first divided into subregions using the LarvalDispersal vector and then further divided evenly by sex of each subregion.
11. **Recruitment Function Type (required)**. This equation will be used to calculate recruitment into each subregion in the area of interest. For a detailed explanation of each equation, please refer to the *Recruitment* section. Each equation requires a different set of recruitment parameters. Be sure that the required parameters for the desired equation are included.
12. **Spawners by Individuals or Weight**. Specifies whether the spawner abundance used in the recruitment function should be calculated in terms of number of individuals or in terms of biomass (weight). If 'Weight' is selected, the user must provide a 'Weight' vector alongside the survival matrix in the Population Parameters CSV File. The 'Alpha' and 'Beta' parameters provided by the user should correspond to the selected choice.
13. **Alpha**. Specifies the shape of the stock-recruit curve. Used only for the *Beverton-Holt* and *Ricker* recruitment functions.
14. **Beta**. Specifies the shape of the stock-recruit curve. Used only for the *Beverton-Holt* and *Ricker* recruitment functions.
15. **Recurring Number of Recruits**. Specifies the total number of recruits that come into the population at each time step (a fixed number). Used only for the *Fixed Recruitment* function.

Migration Parameters

For a species which migrates, this option will include source/sink population dynamics in the model. The migration is done on a class basis, so there is opportunity for each age/stage to have separate migratory patterns.

16. **Migration Matrix CSV Folder (optional)**. If migration is checked, the selected folder should contain CSV migration matrices to be used in the simulation. Each CSV file contains a single migration matrix corresponding to the age/stage that migrates. Not all ages/stages require migration matrices, only those ages/stages that migrate.

    Naming Conventions: Any alphanumeric string of characters. Best to avoid whitespace characters.

    Example Filepath: \InVEST\Fisheries\Inputs_Lobster\Migrations\

    Migration Matrix CSV Files. For each age/stage where migration occurs, there should be a single CSV within the migration directory. The name of the CSV can be anything, but **MUST** end with an underscore followed by the name of the age or stage. This **MUST** correspond to an age or stage within the Population Parameters CSV File. For migration from the 'adult' class for example, a migration file might be named 'migration_adult.csv'. The CSV should contain nothing besides subregion names and migration values. The first row and first column should be the names of the subregions in the Population Parameters CSV File, listed in the same order. The columns represent the sources — the subregions **FROM** which the migration occurs; each column should therefore sum to 1. The rows represent the sinks — the subregions **TO** which the migration occurs. The cells within the matrix should be a DECIMAL REPRESENTATION of percentage of the source's population which will migrate to the sink.

    Naming Convetions: Any alphanumeric string of characters. Best to avoid whitespace characters. Must end with the age/stage name, such as '_ageName.csv'

    Filetype: Comma Separated Values (CSV)

    Example Filepath: \InVEST\Fisheries\Inputs_Lobster\Migrations\migration_adult.csv

    Example Migration CSV File

Migration	Subregion_1 (Src)	Subregion_2	...	Subregion_N
Subregion_1 (Sink)	<float>	<float>	...	<float>
Subregion_2	<float>	<float>	...	<float>
...	...	...	...	...
Subregion_N	<float>	<float>	...	<float>

Valuation Parameters

17. **Fraction of Harvest Kept After Processing (required)**. This is the decimal representation of the percentage of harvested catch remaining after post-harvest processing is complete. (Either by weight or by number of individuals, as set in the Recruitment Parameters)

18. **Unit Price (required)**. Specifies the price per harvest unit. Valuation is intended to give a rough idea of the current market value for an equilibrated population based on user-defined price parameters. If 'Harvest by Individuals or Weight' was set to 'Individuals', this should be the price per individual. If set to 'Weight', this should be the price per unit weight. Weight units should agree with the units implied by the Weight column of the Population Parameters CSV file.

Habitat Scenario Tool

The goal of the Habitat Scenario Tool is to calculate new survival rates from natural mortality of a baseline population given the dependencies of certain classes on certain habitats and the change in area of those habitats over certain subregions.

Upon opening the Habitat Scenario Tool, the user is presented with an interface containing a set of parameters through which to submit inputs. Information about each parameter is provided below. Once the user has entered all necessary inputs, the user can start the model run by pressing 'Run'. If any errors occur, InVEST will stop the model run and provide feedback to the user about what caused the error through a message screen.

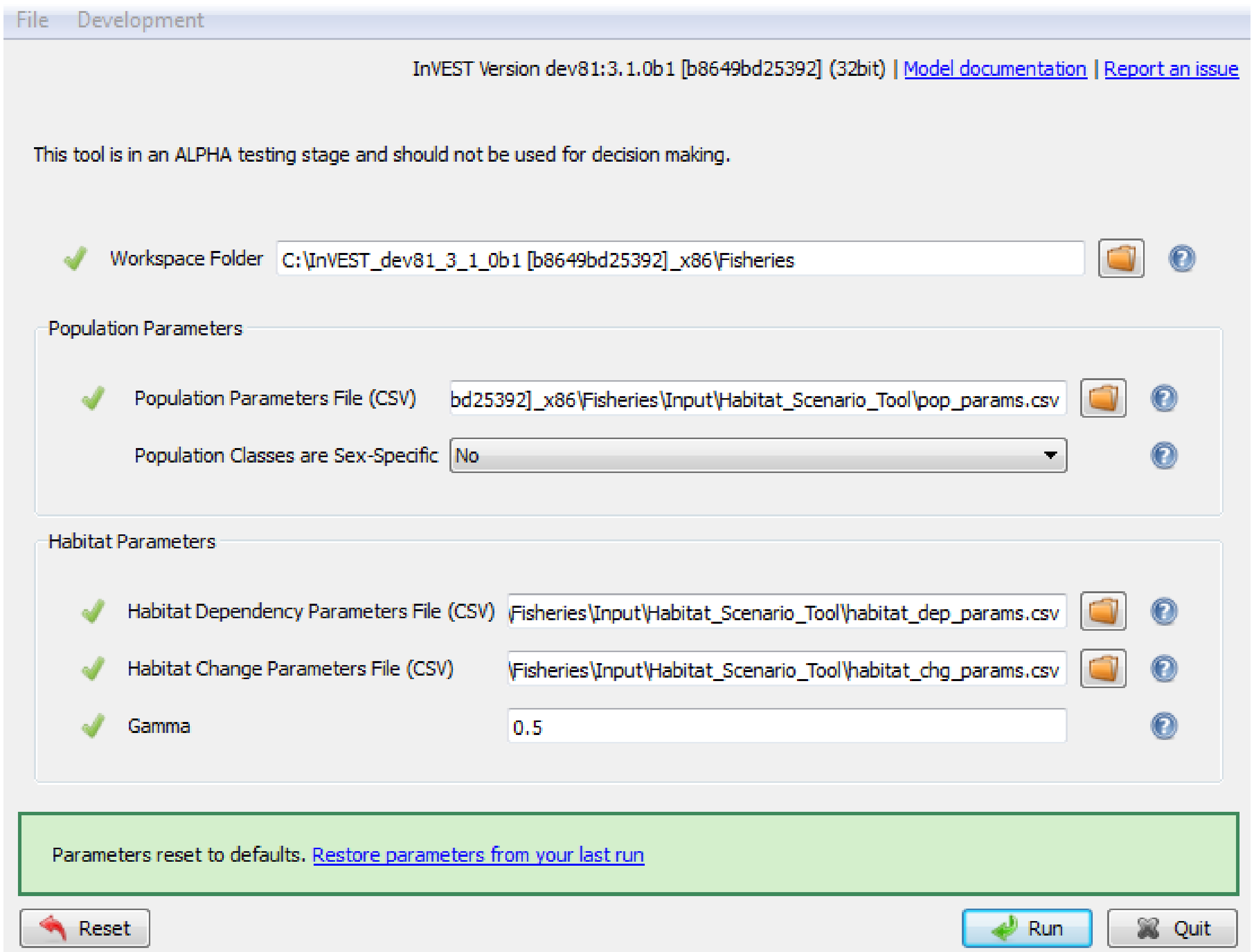


Fig. 5.47: Example User Interface for Habitat Scenario Tool

General Parameters

1. **Workspace (required)**. The selected folder is used as the workspace where all intermediate and output files will be written. If the selected folder does not exist, it will be created. If datasets already exist in the selected folder, they will be overwritten.

   Naming Conventions: Any alphanumeric string of characters. Best to avoid whitespace characters.

   Example Filepath: \InVEST\Fisheries\

Population Parameters

2. **Population Parameters File (CSV) (required)**. The provided CSV file should contain all necessary parameters for population classes based on age/stage, sex, and subregion - excluding possible migration parameters. See the Population Parameters CSV File description in the *Core Model* section for information about the file format.

   Naming Conventions: Any alphanumeric string of characters. Best to avoid whitespace characters.

   Filetype: Comma Separated Values (CSV)

   Example Filepath: \InVEST\Fisheries\Input\Habitat_Scenario_Tool\pop_params.csv

3. **Population Classes are Sex-Specific (required)**. Specifies whether or not the population classes provided in the Population Parameters CSV File are distinguished by sex.

Habitat Parameters

4. **Habitat Dependency Parameters File (CSV)**. The provided CSV file should contain the habitat dependencies (0-1) for each age/stage for each habitat type that is also provided in the Habitat Change CSV File.

   Naming Conventions: Any alphanumeric string of characters. Best to avoid whitespace characters.

   Filetype: Comma Separated Values (CSV)

   Example Filepath: \InVEST\Fisheries\Input\Habitat_Scenario_Tool\habitat_dep_params.csv

   Example Habitat Dependency Parameters CSV File

Habitats	Class_1	Class_2	...	Class_N
Habitat_1	<float>	<float>	...	<float>
Habitat_2	<float>	<float>	...	<float>
...	...	...	...	...
Habitat_N	<float>	<float>	...	<float>

5. **Habitat Area Change File (CSV)**. The provided CSV file should contain the percent changes in habitat area by subregion (if applicable). The habitats included should be those that the population depends upon at any age/stage.

   Name: Any alphanumeric string, avoid spaces.

   Filetype: Comma Separated Values (CSV)

   Example Filepath: \InVEST\Fisheries\Input\Habitat_Scenario_Tool\habitat_chg_params.csv

   Example Habitat Area Change File (CSV)

Habitats	Subregion_1	Subregion_2	...	Subregion_N
Habitat_1	<float>	<float>	...	<float>
Habitat_2	<float>	<float>	...	<float>
...	...	...	...	...
Habitat_N	<float>	<float>	...	<float>

6. **Gamma**. Describes the relationship between the change in habitat area and a change in survival of age/stage dependent on that habitat. Specify a value between 0 and 1.

5.14.9 Interpreting Results

Core Model Results

Upon successful completion of the model, the workspace folder will contain 'intermediate' and 'output' sub-folders. These two folders hold the data generated by the model. Most users will primarily be interested in data contained within the 'output' folder.

Intermediate Outputs Folder

The intermediate folder contains information used for final calculations. Intermediate outputs provide a more comprehensive look at how the final outputs were generated.

1. **Population Breakdown**. The output CSV file details the number of individuals within each class in each subregion, for every time step.

   Example Filepath: \intermediate\<pop_params_name>_population_by_time_step.csv

Final Outputs Folder

1. **HTML Summary of Results**. A page which displays the final harvest after equilibration, and the cumulative harvest across the entire area of interest per time step up to the equilibrated time step. The second table, 'Final Harvest by Subregion After XX Time Steps', shows the final harvest (by individuals or weight, depending on inputs) for each subregion. If valuation of the harvest was selected in the inputs, this will also include a column for the valuation of each subregion harvest (in the input currency). The bottom table, 'Time Step Breakdown', shows the cumulative harvest across all subregions for each time step before the model equilibrates. If valuation of the harvest was selected in the inputs, this will also include a column for valuation of the subregion harvest using the input currency. The 'Equilibrated?' column indicates whether the model reached equilibrium for each given time step (N=No, Y=Yes).

   Example Filepath: \output\<pop_params_name>_results_page.html

2. **CSV Summary of Results**. A summary of results, formatted as a CSV file.

   Example Filepath: \output\<pop_param_name>_results_table.csv

3. **Modified AOI**. A copy of the AOI layer, but with either one or two additional attributes. The layer will always contain 'Hrv_Total', which is the final harvest (in either number of individuals or weight, depending on inputs) within that subregion at the final time step. Additionally, if valuation was checked, the layer will also contain 'Val_Total', the total value of the harvest within that subregion using the currency from the inputs.

   Example Filepath: \output\<aoi_name>_results_aoi.shp

Habitat Scenario Tool Results

Upon completion of a successful model run, the workspace folder will contain an 'output' sub-folder. No intermediate files are created.

Final Outputs Folder

1. **Modified Population Parameters CSV File** A new population parameters file with an adjusted survival matrix based on the Habitat Scenario equation.

    Example Filepath: \output\<pop_params>_modified.csv

5.14.10 References

Arkema, K.K., G. Verutes, S.A. Wood, C. Clarke, S. Rosado, M. Canto, A. Rosenthal, M. Ruckelshaus, G. Guannel, J. Toft, J. Fariesa, J.M. Silver, R. Griffin, A.D. Guerry. Improving the margins: Modeling ecosystem services leads to better coastal plans. *In review*

Baker, R., P. Levin, and T. Minello. 2008. The link between coastal wetlands and white shrimp fishery production in the northern Gulf of Mexico. ICES CM 2008/M:11

Clarke, C., M. Canto, and S. Rosado. 2013. Belize Integrated Coastal Zone Management Plan. Coastal Zone Management Authority and Institute, Belize City. 423 p.

Food and Agriculture Organization of the United Nations. 2014a. The State of World Fisheries and Aquaculture 2014: Opportunities and Challenges. Food and Agriculture Organization of the United Nations. Rome

Guannel, G., A. Guerry, J. Brenner, J. Faries, M. Thompson, J. Silver, R. Griffin, J. Proft, M. Carey, J. Toft, G. Verutes. 2014. Changes in the delivery of ecosystem services in Galveston Bay, TX, under a sea-level rise scenario. Report available upon request.

Higgins, K., A. Hastings, J.N. Sarvela, and L.W. Botsford. 1997. Stochastic dynamics and deterministic skeletons: population behavior of Dungeness crab. Science 276: 1431-1435

Puga, R., S. Hernández, J. López and M.E. de León. 2005. Bioeconomic modeling and risk assessment of the Cuban fishery for spiny lobster *Panulirus argus*. Fisheries Research 75: 149–163

Toft, J.E., J.L. Burke, M.P. Carey, C.K. Kim, M. Marsik, D.A. Sutherland, K.K. Arkema, A.D. Guerry, P.S. Levin, T.J. Minello, M. Plummer, M.H. Ruckelshaus and H.M. Townsend. 2013. From mountains to sound: modelling the sensitivity of Dungeness crab and Pacific oyster to land-sea interactions in Hood Canal, WA. ICES J. Mar. Sci. 71(3): 725-738

5.15 Crop Production

5.15.1 Summary

NOTE THE CROP PRODUCTION TOOL IS IN A ALPHA/UNSTABLE STATE. IT SHOULD NOT BE USED IN DECISION-MAKING REPORTS UNTIL IT IS RELEASED AS A STABLE MODEL.

5.15.2 Introduction

Expanding agricultural production and closing yield gaps is a key strategy for many governments and development agencies focused on poverty alleviation and achieving food security. However, conversion of natural habitats to agri-

cultural production sites impacts other ecosystem services that are key to sustaining the economic benefits that agriculture provides to local communities. Intensive agricultural practices can add to pollution loads in water sources, often necessitating future costly water purification methods. Overuse of water also threatens the supply available for hydropower or other services. Still, crop production is essential to human well-being and livelihoods. The InVEST crop production model allows detailed examination of the costs and benefits of this vital human enterprise, allowing exploration of questions such as:

- How would different arrangement or selection of cropping systems compare to current systems in terms of total production? Could switching crops yield higher economic returns or nutritional value?
- What are the impacts of crop intensification on ecosystem services? If less land is used to produce equal amounts of food by increasing intensification, is the net result on ecosystem services production positive or negative?
- How can we evaluate different strategies for meeting increasing food demand while minimizing impact on ecosystem services?

5.15.3 The Model

The InVEST crop production model will produce estimates of crop yield, from existing data, percentile summaries, and modeled predictions. For existing or modeled crop yields, the model can also generate estimates of crop value.

Observed data: The crop yield model supplies observed yields, based on FAO and sub-national datasets for 175 crops, as tons/ha (Monfreda et al. 2008). If a crop type submitted by the user is not grown in that region, the model will not return a value for those pixels; crops can be moved around within a region in which they are grown, but novel cropping systems cannot be introduced in minimum mode. The model will also return existing inputs for that crop (in that region), as percent of land irrigated (for 15 crops for which there are data), and amount of N, P, and K applied/ha (for 140 crops for which there are data). The model can provide nutrition information for all crops and economic production if additional cost information is provided for fertilizer, nutrients, labor, seed, and machinery (this information is already included in the model for 12 staple crops in 2012: barley, maize, oil palm, potato, rapeseed, rice, rye, soybean, sugar beet, sugar cane, sunflower, and wheat).

Percentile summaries: This option allow the user to explore yields under different management scenarios, picking from a range of “intensification” levels. The user can supply a percentile raster, with each pixel of agricultural land coded 25, 50, 75, or 95 as a proxy for low, medium, high, or maximum productivity, presumably achieved under corresponding degrees of intensification. This will return the 25th, 50th, 75th, or 95th percentile yields, respectively, for the crop of interest in all regions of similar climate and income. The user also has an option to select a World Bank Income

Classification to explore different productivity levels possible under different income conditions, with accompanying assumptions about how this would change their access to technology or other capital. This feature is optional and should be interpreted with caution (i.e., the feasibility of moving from one income category to another), and if the user does not make a selection the model will use the actual income classification for that region. Because there is no information about the actual quantities of fertilizer and irrigation needed to achieve these different percentiles of production, this option does not allow economic valuation of production, but can provide nutrition production information if desired.

Modeled yields: For 12 staple crops for which yields have been modeled globally by Mueller et al. (2011), the model can provide estimates of both yields and inputs (fertilizer and irrigation), in the same units as above. These crops include barley, maize, oil palm, potato, rapeseed, rice, rye, soybean, sugar beet, sugar cane, sunflower, and wheat. To run this model, the user must provide rasters of nitrogen, phosphate, and potash application rate (kg/ha) and an irrigation raster (0 for pixels that are not irrigated and 1 for pixels that are) that cover all cropped areas of interest. The model returns crop yields and economic and nutritional value.

The crop value model can use the yields and/or inputs generated by the yield model, or can be run with yield maps derived from other models or data sources (e.g., SSURGO). Crop yields can be valued in terms of economic returns or in terms of nutrition. To calculate economic returns, the model requires yield maps, as well as maps of fertilizer and irrigation rates corresponding to those yields, and combines this information with crop price and cost datasets to calculate the total expected returns (yields x area x price – inputs x input costs – area x other costs). To calculate nutrition, the model only requires yield maps of all food crops produced, and the user can select from 33 macro and micronutrients to map or summarize. This model can be combined with our nutrition demand model, which multiplies population density by recommended daily allowances of the same nutrients, to determine what proportion of nutritional requirements can be met from local food production.

How it Works

Calculating Yield and Production

Method 1: Observed Regional Yields (Observed)

$$ProductionPerCell_{crop,x,y} = ObservedLocalYieldPerHectare_{crop,x,y} * HectaresPerCell$$

$$ProductionTotal_{crop} = \sum_{x,y} ProductionPerCell_{crop,x,y}$$

Method 2: Climate-specific Distribution of Observed Yields (Percentile)

$$YieldPerHectare_{crop,percentile,x,y} = (ObservedClimateBinYield_{crop,precentile,climatebin} \mid ClimateBin_{x,y})$$

$$ProductionPerCell_{crop,percentile,x,y} = YieldPerHectare_{crop,percentile,x,y} * HectaresPerCell$$

$$ProductionTotal_{crop,percentile} = \sum_{x,y} ProductionPerCell_{crop,percentile,x,y}$$

Method 3: Yield Regression Model with Climate-specific Parameters (Modeled)

$$PercentMaxYieldNitrogen_{x,y} = \left(1 - Bnp_{crop,climatebin} * e^{-Cn_{crop,climatebin} * NitrogenAppRate_{x,y}} \mid ClimateBin_{x,y}\right)$$

$$PercentMaxYieldPhosphorus_{x,y} = \left(1 - Bnp_{crop,climatebin} * e^{-Cp_{crop,climatebin} * PhosphorusAppRate_{x,y}} \mid ClimateBin_{x,y}\right)$$

$$PercentMaxYieldPotassium_{x,y} = \left(1 - Bk_{crop,climatebin} * e^{-Ck_{crop,climatebin} * PotassiumAppRate_{x,y}} \mid ClimateBin_{x,y}\right)$$

$$MaxYieldNitrogen_{x,y} = MaxYield_{crop,climatebin} * PercentMaxYieldNitrogen_{x,y}$$

$$MaxYieldPhosphorus_{x,y} = MaxYield_{crop,climatebin} * PercentMaxYieldPhosphorus_{x,y}$$

$$MaxYieldPotassium_{x,y} = MaxYield_{crop,climatebin} * PercentMaxYieldPotassium_{x,y}$$

$$YieldPerHectare_{crop,x,y} = \begin{cases} min\left(MaxYieldNitrogen, MaxYieldPhosphorus, MaxYieldPotassium\right) \\ min\left(MaxYieldNitrogen, MaxYieldPhosphorus, MaxYieldPotassium, MaxYieldRain\right. \end{cases}$$

$ProductionPerCell_{crop,x,y} = YieldPerHectare_{crop,x,y} * HectaresPerCell_{x,y}$

$ProductionTotal_{crop} = \sum_{x,y} ProductionPerCell_{crop,x,y}$

Calculating Nutritional Contents from Production

$NutrientAmount_{crop,nutrient} = NutrientAmountPerTonCrop_{crop,nutrient} * ProductionTotal_{crop} * (1 - FractionRefuse)$

$NutrientAmountTotal_{nutrient} = \sum_{crops} NutrientAmount_{crop,nutrient}$

Calculating Economic Returns

$KilogramInputTotalCosts_{crop,x,y} = \sum_{fertilizer} (FertKgPerHectare_{fertilizer,x,y} * CostPerKg_{crop,fertilizer} * Hectares$

$HectareInputTotalCosts_{crop,x,y} = \sum_{inputs} CostPerHectare_{input,x,y} * HectaresPerCell$

$Cost_{crop,x,y} = KilogramInputTotalCosts_{crop,x,y} + HectareInputTotalCosts_{crop,x,y}$

$Revenue_{crop,x,y} = Production_{crop,x,y} * Price_{crop}$

$Returns_{crop,x,y} = Revenue_{crop,x,y} - Cost_{crop,x,y}$

$ReturnsTotal_{crop} = \sum_{x,y} Returns_{crop,x,y}$

Limitations and Simplifications

The current version of the model is a coarse global model driven mostly by climate and optionally by management. This model is therefore not able to capture the variation in productivity that occurs across heterogeneous landscapes. A rocky hill slope and a fertile river valley, if they share the same climate, would be assigned the same yield in the current model. This is a problem if the question of interest is where: where to prioritize future habitat conversion; or where farming is most productive and least destructive.

Spatial downscaling of the current coarse global model is necessary to make the crop model more useful in local land-use decisions. Our approach will be to acquire local yield data that can be compared to the regression model yields to determine where the model is overestimating yields and where it is underestimating. The resulting differences can be related to other variables such as slope, aspect, elevation, soil fertility, and soil depth, and any significant relationships can be used to refine the current model. The coarse model will still be used to arrive at the general magnitude of yield for a given climate and intensification level, and the finer-scale differences will essentially tune the coarse model up or down. To do this we need:

- Field-level (or better) yield data across a wide representation of soils, topographies and climates
- Soil and topographic data at the same level of resolution as the yield data

If you have or intend to take such data and are interested in collaborating with us, please contact Becky Chaplin-Kramer at bchaplin@stanford.edu

5.15.4 Data Needs

- Monfreda Dataset
  - Observed Crop Yields (provided in units of tons per harvested hectare (growing season))
- Foley Lab Datasets
  - Climate-bin Based Tables

– Regression Model Parameters

5.15.5 Running the Model

General Parameters

1. **Workspace Folder** The selected folder is used as the workspace where all intermediate and final output files will be written. If the selected folder does not exist, it will be created. If datasets already exist in the selected folder, they will be overwritten.

2. **Results Suffix (Optional)** This text will be appended to the end of the output folders to help separate outputs from multiple runs. Please see the *Interpreting Results* section for an example folder structure for outputs.

3. **LULC Lookup Table (CSV)** A CSV table used to manage the relationship between the lulc codes and the crop dataset. The provided CSV file should contain a table with two columns: a 'lulc-class' column, a 'code' column, and a 'is_crop' column. The 'lulc-class' column contains the names of each lulc-class used in the model, the 'code' column contains the associated code used to represent that lulc-class in the LULC Map, and the 'is_crop' column contains a boolean value indicating whether the given lulc-class is a crop. If 'is_crop' is set to True, the Spatial Dataset must contain tables and maps associated with that crop. Any non-negative integer value can be used as a 'code' value.

lulc-class	code	is_crop
other	0	False
maize	1	True
soybean	2	True
rice	3	True
...	...	

4. **LULC Map (Raster)** A GDAL-supported raster representing a crop management scenario. Each cell value in the raster should be a valid integer code that corresponds to a lulc-class in the LULC Lookup Table file. The NoData value should be set to a number not existing in the LULC Lookup Table.

int	int
int	int

5. **Fertilizer Application Rate Maps (Rasters)** A set of GDAL-supported rasters representing the amount of Nitrogen (N), Phosphorus (P2O5), and Potash (K2O) applied to each area of land. These maps are required for the regression model yield function and are an optional input for all yield functions when calculating economic returns. Each cell value in the raster should be a non-negative float value representing the amount of fertilizer applied in units of kilograms per hectare (kgs/ha). Each file should be prepended with the name of the fertilizer (nitrogen, phosphorus, potash) in lowercase, followed by an underscore to help the program search for the matching file. The Fertilizer Maps should have the same dimensions and projection as the provided LULC Map.

float	float
float	float

Folder Structure

```
.
|-- fertilizer_maps_folder
    |-- nitrogen_application_map.tif
    |-- phosphorus_application_map.tif
    |-- potash_application_map.tif
```

6. **Crop Production Model Spatial Dataset Folder**

Folder Structure

```
.
|-- spatial_dataset_folder
    |-- climate_bin_maps
    |   |-- [crop]_climate_bin_map (*.tif)
    |-- climate_percentile_yield
    |   |-- [crop]_percentile_yield_table.csv
    |-- climate_regression_yield
    |   |-- [crop]_regression_yield_table.csv
    |-- observed_yield
        |-- [crop]_yield_map (*.tif)
```

Embedded Data for Functions Based on Climate (Percentile and Regression Functions)

Crop Climate-Bin Maps (Rasters) A set of GDAL-supported rasters representing the climate-bin that a given area of land is located within for each particular crop. Each raster contains a set of values between 0 and 100. Zero-values represent areas that do not exist within a climate-bin, such as an ocean. Values 1 through 100 correspond to a particular climate-bin. The climate-bin maps reside in the 'climate_bin_maps' folder of the provided spatial dataset.

int	int
int	int

Embedded Data for Observed Regional Yields

Observed Crop Yield Maps (Rasters) A set of GDAL-supported rasters representing the observed regional crop yield. Each cell value in the raster should be a non-negative float value representing the amount of crop produced in units of tons per hectare (tons/hectare). The observed yield maps reside in the 'observed_yield' folder of the provided spatial dataset.

float	float
float	float

Embedded Data for Climate-specific Distribution of Observed Yields

Percentile Yield Table (CSV) The provided CSV tables should contain information about the average crop yield occurring within each climate-bin across several income levels for each crop in units of tons per hectare (tons/ha). The table must have a 'climate_bin' column containing values 0 through 100. The table must have at least one additional column representing a percentile yield within the given climate-bin for a particular crop - an example set of columns could be: 'yield_25th', 'yield_50th', 'yield_75th', 'yield_95th'. So, this example table would have the following columns: 'crop', 'climate_bin', 'yield_25th', 'yield_50th', 'yield_75th', 'yield_95th'. Each file should be prepended with the name of the crop in lowercase, followed by an underscore to help the program parse the file. The tables reside in the 'climate_percentile_yield' folder of the provided spatial dataset.

climate_bin	yield_25th	yield_50th	yield_75th	yield_95th	...
1	<float>	<float>	<float>	<float>	...
2	<float>	<float>	<float>	<float>	...
3	<float>	<float>	<float>	<float>	...
...	...	...	...	...	...

e.g. 'maize_percentile_yield_table.csv'

Embedded Data for Yield Regression Model with Climate-specific Parameters

Regression Model Yield Table (CSV) The provided CSV tables should contain information useful for calculating the yield of a crop located in a particular climate-bin based on the limiting factor. The table must have the following columns: 'climate_bin', 'yield_ceiling', 'yield_ceiling_rf', 'b_nut', 'b_K2O', 'c_N', 'c_P2O5', and 'c_K2O'. Each file should be prepended with the name of the crop in lowercase, followed by an underscore to help the program search for the matching file. Currently, the regression model yield function is useful to a small subset of the crops provided in the dataset. The tables reside in the 'climate_regression_yield' folder of the provided spatial dataset.

climate_bin	yield_ceiling	yield_ceiling_rf	b_nut	b_K2O	c_N	c_P2O5	c_K2O
1	<float>	<float>	<float>	<float>	<float>	<float>	<float>
2	<float>	<float>	<float>	<float>	<float>	<float>	<float>
3	<float>	<float>	<float>	<float>	<float>	<float>	<float>
...	...	...	...	...	...	...	...

e.g. 'maize_regression_yield_table.csv'

Parameters for Yield Regression Model with Climate-specific Parameters

7. **Irrigation Map (Raster)** A GDAL-supported raster representing whether irrigation occurs or not. A zero value indicates that no irrigation occurs. A one value indicates that irrigation occurs. The Irrigation Map should have the same dimensions and projection as the provided LULC Map.

int	int
int	int

Note: The regression yield function also requires the 'Fertilizer Application Rate Maps' as an input.

Parameters for Calculating Nutritional Contents from Production

8. **Nutrient Contents Table (CSV)** A CSV table containing information about the nutrient contents of each crop. The values provided are assumed to be given in relation to one ton of harvest crop biomass. The 'crop' and 'fraction_refuse' columns must be provided in the table. The 'fraction_refuse' column is expected to contain a value between 0 and 1 representing the fraction of the harvested crop that is considered refuse and does not contain nutritional value.

crop	fraction_refuse	protein	lipid	energy	ca	ph	...
maize	<float>	<float>	<float>	<float>	<float>	<float>	...
soybean	<float>	<float>	<float>	<float>	<float>	<float>	...
...	...	...	...	...	...	...	...

Parameters for Calculating Economic Returns

9. **Economics Table (CSV)** A CSV table containing information related to market price of a given crop and the costs involved with producing that crop.

crop	price_per_ton	cost_nitrogen_per_kg	cost_phosphorous_per_kg	cost_potash_per_kg	cost_labor_per_ha	cost_machine_per_ha	cost_seed_per_ha	cost_irrigation_per_ha
maize	<float>	<float>	<float>	<float>	<float>	<float>	<float>	<float>
soybean	<float>	<float>	<float>	<float>	<float>	<float>	<float>	<float>
...	...	...	...	...	...	...	...	...

5.15.6 Interpreting Results

Outputs Folder Structure

A unique set of outputs shall be created for each yield function that is run such that the folder structure may look as follows:

```
.
|-- outputs
    |-- climate_percentile_yield_[results suffix]
    |   |-- results_table (.csv)
    |   |-- yield_map (.tif)
```

```
    |   |-- production_map (.tif)
    |   |-- cost_map (.tif)
    |   |-- revenue_map (.tif)
    |   |-- returns_map (.tif)
    |-- climate_regression_yield_[results suffix]
    |   |-- results_table (.csv)
    |   |-- yield_map (.tif)
    |   |-- production_map (.tif)
    |   |-- cost_map (.tif)
    |   |-- revenue_map (.tif)
    |   |-- returns_map (.tif)
    |-- observed_yield_[results suffix]
        |-- results_table (.csv)
        |-- yield_map (.tif)
        |-- production_map (.tif)
        |-- cost_map (.tif)
        |-- revenue_map (.tif)
        |-- returns_map (.tif)
```

Outputs

1. **Results Table (CSV)**

crop	production	(percentile)	(return)	(revenue)	(cost)	(nutrient_a)	(nutrient_b)	(etc.)
maize	<float>	<str>	<float>	<float>	<float>	<float>	<float>	...
soybean	<float>	<str>	<float>	<float>	<float>	<float>	<float>	...
...	...	...	...	...	...	...	...	...

2. **Crop Yield Map (Raster)** A set of GDAL-supported rasters spatially representing the per-hectare yield for a given crop in each cell. Each cell value in the raster shall be a non-negative float value representing the yield area under the given scenario in units of tons per hectare (tons/ha).

float	float
float	float

3. **Crop Production Map (Raster)** A GDAL-supported raster spatially representing the total production for a given crop in each cell. Each cell value in the raster shall be a non-negative float value representing the total production over the cell's area under the given scenario in units of tons.

float	float
float	float

4. **Economic Cost Map (Raster) (Optional)** A GDAL-supported raster representing the economic cost associated with the crops. Each cell value in the raster shall be a float value representing the cost generated under the given scenario in units of the currency from the user-provided Economics Table. If insufficient data is provided within a given cell, the cell will contain a NoData value.

float	float
float	float

5. **Economic Revenue Map (Raster) (Optional)** A GDAL-supported raster representing the economic revenue generated by the crops. Each cell value in the raster shall be a float value representing the revenue generated under the given scenario in units of the currency from the user-provided Economics Table. If insufficient data is provided within a given cell, the cell will contain a NoData value.

float	float
float	float

6. **Economic Returns Map (Raster) (Optional)** A GDAL-supported raster representing the economic returns generated by the crops. Each cell value in the raster shall be a float value representing the return (revenue minus cost) generated under the given scenario in units of the currency from the user-provided Economics Table. If insufficient data is provided within a given cell, the cell will contain a NoData value.

float	float
float	float

5.15.7 References

Monfreda et al. 2008

Mueller et al. 2012

5.15.8 Appendix I

Available Crop Data within Global Dataset

Crop	Observed Model	Percentile Model	Regression Model
Abaca	Yes	Yes	No
Agave	Yes	Yes	No
Alfalfa	Yes	Yes	No
Almond	Yes	Yes	No
Aniseetc	Yes	Yes	No
Apple	Yes	Yes	No
Apricot	Yes	Yes	No
Areca	Yes	Yes	No
Artichoke	Yes	Yes	No
Asparagus	Yes	Yes	No
Avacado	Yes	Yes	No
Bambara	Yes	Yes	No
Banana	Yes	Yes	No
Barley	Yes	Yes	Yes
Bean	Yes	Yes	No
Beetfor	Yes	Yes	No
Berrynes	Yes	Yes	No
Blueberry	Yes	Yes	No
Brazil	Yes	Yes	No
Broadbean	Yes	Yes	No
Buckwheat	Yes	Yes	No
Cabbage	Yes	Yes	No
Cabbagefor	Yes	Yes	No
Canaryseed	Yes	Yes	No
Carob	Yes	Yes	No
Carrot	Yes	Yes	No
Carrotfor	Yes	Yes	No
Cashew	Yes	Yes	No
Cashewapple	Yes	Yes	No
Cassava	Yes	Yes	No
Castor	Yes	Yes	No

Continued on next page

Table 5.1 – continued from previous page

Cauliflower	Yes	Yes	No
Cerealnes	Yes	Yes	No
Cherry	Yes	Yes	No
Chestnut	Yes	Yes	No
Chickpea	Yes	Yes	No
Chicory	Yes	Yes	No
Chilleetc	Yes	Yes	No
Cinnamon	Yes	Yes	No
Citrusnes	Yes	Yes	No
Clove	Yes	Yes	No
Clover	Yes	Yes	No
Cocoa	Yes	Yes	No
Coconut	Yes	Yes	No
Coffee	Yes	Yes	No
Coir	Yes	No	No
Cotton	Yes	Yes	No
Cowpea	Yes	Yes	No
Cranberry	Yes	Yes	No
Cucumberetc	Yes	Yes	No
Currant	Yes	Yes	No
Date	Yes	Yes	No
Eggplant	Yes	Yes	No
Fibrenes	Yes	Yes	No
Fig	Yes	Yes	No
Flax	Yes	Yes	No
Fonio	Yes	Yes	No
Fornes	Yes	Yes	No
Fruitnes	Yes	Yes	No
Garlic	Yes	Yes	No
Ginger	Yes	Yes	No
Gooseberry	Yes	Yes	No
Grape	Yes	Yes	No
Grapefruitetc	Yes	Yes	No
Grassnes	Yes	Yes	No
Greenbean	Yes	Yes	No
Greenbroadbean	Yes	Yes	No
Greencorn	Yes	Yes	No
Greenonion	Yes	Yes	No
Greenpea	Yes	Yes	No
Groundnut	Yes	Yes	No
Gums	Yes	No	No
Hazelnut	Yes	Yes	No
Hemp	Yes	Yes	No
Hempseed	Yes	Yes	No
Hop	Yes	Yes	No
Jute	Yes	Yes	No
Jutelikefiber	Yes	Yes	No
Kapokfiber	Yes	Yes	No
Kapokseed	Yes	Yes	No
Karite	Yes	Yes	No

Continued on next page

Table 5.1 – continued from previous page

Kiwi	Yes	Yes	No
Kolant	Yes	Yes	No
Legumenes	Yes	Yes	No
Lemonlime	Yes	Yes	No
Lentil	Yes	Yes	No
Lettuce	Yes	Yes	No
Linseed	Yes	Yes	No
Lupin	Yes	Yes	No
Maize	Yes	Yes	Yes
Maizefor	Yes	Yes	No
Mango	Yes	Yes	No
Mate	Yes	Yes	No
Melonetc	Yes	Yes	No
Melonseed	Yes	Yes	No
Millet	Yes	Yes	No
Mixedgrain	Yes	Yes	No
Mixedgrass	Yes	Yes	No
Mushroom	Yes	Yes	No
Mustard	Yes	Yes	No
Nutmeg	Yes	Yes	No
Nutnes	Yes	Yes	No
Oats	Yes	Yes	No
Oilpalm	Yes	Yes	Yes
Oilseedfor	Yes	Yes	No
Oilseednes	Yes	Yes	No
Okra	Yes	Yes	No
Olive	Yes	Yes	No
Onion	Yes	Yes	No
Orange	Yes	Yes	No
Papaya	Yes	Yes	No
Pea	Yes	Yes	No
Peachetc	Yes	Yes	No
Pear	Yes	Yes	No
Pepper	Yes	Yes	No
Peppermint	Yes	Yes	No
Persimmon	Yes	Yes	No
Pigeonpea	Yes	Yes	No
Pimento	Yes	Yes	No
Pineapple	Yes	Yes	No
Pistachio	Yes	Yes	No
Plantain	Yes	Yes	No
Plum	Yes	Yes	No
Popcorn	Yes	Yes	No
Poppy	Yes	Yes	No
Potato	Yes	Yes	Yes
Pulsenes	Yes	Yes	No
Pumpkinetc	Yes	Yes	No
Pyrethrum	Yes	Yes	No
Quince	Yes	Yes	No
Quinoa	Yes	Yes	No

Continued on next page

Table 5.1 – continued from previous page

Ramie	Yes	Yes	No
Rapeseed	Yes	Yes	No
Raspberry	Yes	Yes	No
Rice	Yes	Yes	Yes
Rootnes	Yes	Yes	No
Rubber	Yes	Yes	No
Rye	Yes	Yes	No
Ryefor	Yes	Yes	No
Safflower	Yes	Yes	No
Sesame	Yes	Yes	No
Sisal	Yes	Yes	No
Sorghum	Yes	Yes	No
Sorghumfor	Yes	Yes	No
Soybean	Yes	Yes	Yes
Sourcherry	Yes	Yes	No
Spicenes	Yes	Yes	No
Spinach	Yes	Yes	No
Stonefruitnes	Yes	Yes	No
Strawberry	Yes	Yes	No
Stringbean	Yes	Yes	No
Sugarbeet	Yes	Yes	Yes
Sugarcane	Yes	Yes	Yes
Sugarnes	Yes	Yes	No
Sunflower	Yes	Yes	Yes
Swedefor	Yes	Yes	No
Sweetpotato	Yes	Yes	No
Tangetc	Yes	Yes	No
Taro	Yes	Yes	No
Tea	Yes	Yes	No
Tobacco	Yes	Yes	No
Tomato	Yes	Yes	No
Triticale	Yes	Yes	No
Tropicalnes	Yes	Yes	No
Tung	Yes	Yes	No
Turnipfor	Yes	Yes	No
Vanilla	Yes	Yes	No
Vegetablenes	Yes	Yes	No
Vegfor	Yes	Yes	No
Vetch	Yes	Yes	No
Walnut	Yes	Yes	No
Watermelon	Yes	Yes	No
Wheat	Yes	Yes	Yes
Yam	Yes	Yes	No
Yautia	Yes	Yes	No

Fertilizer Units

Band 1: Kg/ha

Band 2: Precision

- any previous number + .25 = any one of the previous data types but scaling of application rates was maxed out at a doubling when trying to match the FAO consumption

5.15.9 Appendix II - Statistics

Climate Bin Fertilizer

Climate Bin Correlation Coefficient

CHAPTER

SIX

TOOLS TO FACILITATE ECOSYSTEM SERVICE ANALYSES:

6.1 Overlap Analysis Model

6.1.1 Summary

Mapping current uses and summarizing the relative importance of various regions for particular activities is an important first step in marine spatial planning. The InVEST Overlap Analysis Model was designed to produce maps that can be used to identify marine and coastal areas that are most important for human use. Initial development of this model was as two separate models for recreation and fisheries. However, since the underlying approach was fundamentally similar, we combined them into one model that can be used to map not only recreation and fisheries, but also other activities. While this model was envisioned for use in marine areas where space is shared commonly, it is also applicable to locations on land where overlapping uses occur. Inputs include information about where human use activities occur (required), weights that reflect the relative importance of different human uses (optional) and information on spatial variability within uses (optional). Because it simply maps current uses and does not model behavior, this model is not well-suited to the evaluation of how human uses may change in response to changes in the coastal and marine environment. However, it can be used to model scenarios that reflect changes in the areas used by different activities or changes in attributes such as total landings or number of trips that are used to weight activities.

6.1.2 Introduction

Understanding where and how humans use coastal and marine environments is an essential component of marine resource planning and management. Marine and coastal ecosystems are essential places for a variety of activities including fishing (commercial, recreational, subsistence and ceremonial) and recreation (e.g. boating, kayaking, diving, whale-watching). When siting new activities and infrastructure or zoning areas for particular uses, a key step is the identification and visualization of the variety of human uses that occur in the region and the places in which they overlap (e.g., GBRMPA 2003, CDFG 2008, Beck et al. 2009, CRMC 2010). This allows for the identification of hotspots of human use and highlights regions where the compatibility of various activities should be investigated.

The InVEST Overlap Analysis model provides users with a simple framework for mapping and identifying important areas for human use in the marine environment. The model also allows users to include information about a variety of uses of the coastal and marine environment (e.g., commercial fishery logbooks or landings reports data, participation numbers for recreational activities) that can be used to weigh the relative importance of different uses and locations. The model is simple to use, quick to run, and can be applied in any region of the world where there is spatially-explicit information on human uses. The model does not value ecosystem services or estimate the economic value of human uses, but the outputs can be used to identify areas and different user groups that may be affected by policy change. The model produces a map of hotspots for human activities (e.g., fishing activity/fishing grounds) across as many human uses as the user chooses to include. Throughout this chapter we will give examples for both recreation and fisheries.

Using the tool across various categories of human uses may make sense in some instances, but devising schemes for weights will likely be difficult. Outputs can be used to help decision-makers weigh potential conflicts between sectors of spatially-explicit management options that may involve new activities or infrastructure.

The Overlap Analysis Model complements the more involved InVEST recreation model and the fisheries model that is in development. The InVEST Marine Fish Aquaculture model is appropriate for use with single species or groups of species and is used to estimate the quantity and value of fish harvested by commercial fisheries. Additionally, a recreational submodel can be used to predict the amount of recreational fishing effort required to catch the quantity of fish output from the InVEST Marine Fish Aquaculture model. Future more advanced fisheries models will include functionality to incorporate impacts of biogenic habitat on the survival and fecundity of different life-stages of target species, and the ability to wrap around outputs from more complex food-web models (e.g., Ecopath with Ecosim and Atlantis).

6.1.3 The Model

The InVEST Overlap Analysis model was designed to identify marine and coastal areas most important for human use. The model combines the different input layers of human use and computes an "Importance Score" for each grid cell or management area. If users only know where activities occur but do not have additional information to weight the relative importance of different activities, the default model computes an "Importance Score" by summing the number of activities that occur in any particular cell or zone. Although not required to run the model, users can input qualitative (e.g., indices, scores) or quantitative (e.g., catches, effort levels, revenues, profits) information to weight the importance of different locations for an individual activity and to weight activities compared to one another. The model also allows users to down-weight areas or zones used for different activities as a function of their distance from important land-based hubs such as ports, marinas, or public access points. Model outputs are mapped in the coastal region of interest over the specified seascape or management zones. The default model map output is a shapefile showing the frequency of occurrence of activities across the area of interest. If additional weighting information is included, the model also produces a shapefile showing the gradation of importance across cells or zones. The resulting maps can then be used to evaluate the relative importance of different areas in the seascape for the set of human activities included in the analysis. See *Appendix A* for suggestions for data sources.

How it Works

Calculating Frequencies (Model Default)

Users input maps of the locations of multiple human activities. Data is input in a vector format as polygons or points; vector data are rasterized after they are input. In the simplest (default) model, all activities and locations are weighted equally and the model calculates an Importance Score (IS), which is a count of how many activities take place in each grid cell or management zone i:

$$IS_i = \sum_{i,j} U_{ij} I_j \tag{6.1}$$

where U_{ij} = usage of activity j in grid cell or management zone i. U_{ij} is scored by the presence (U_{ij} = 1) or absence (U_{ij} = 0) of the activity in the cell or zone.

Including Weights (Optional)

Users are also given the option to apply different weights to each activity. The two ways in which users can provide these weights are as inter- or intra-activity weights:

1. Inter-activity weight: this allows users to weight the importance of activities relative to one another. Users may choose to give more weight in the analysis to certain activities (e.g., those that generate the highest profits of all fleets in the analysis, or are key employers in the region) and less to other activities. For example, if the user

is examining 3 activities (1. commercial salmon fishing, 2. commercial crab fishing, and 3. commercial kelp harvest) and commercial salmon fishing is deemed to be twice as important as either commercial crab fishing or commercial kelp harvest, then the user would provide weights of (commercial salmon fishing, commercial crab fishing, commercial kelp harvest)= (2,1,1). Inter-activity weights are included in an input .csv table (see "Running the Model" section below); and/or

2. Intra-activity weight: Spatially explicit information about the relative importance of various locations (points or polygons on the map) for a particular activity can be used to weight the scores used in the model calculations. Importance can be measured several ways. For fisheries, weights might be informed by the amount of fish caught or landed, profits earned, safety or accessibility of the fishing ground, or the cultural value of the area. For recreation, they might be determined by the number of visitors or trips to different areas. For example, if the user is examining three commercial harvesting activities and has catch data for each polygon representing those activities, these intra-activity weights can be included by adding a column to the shapefile attribute table of each input activity layer. The name of this column should have no spaces, and this column name will need to be given as an input so that the model knows where these weights are stored.

If intra- or inter-activity weights are included, *IS* is weighted by the importance of the cell (or zone) relative to other cells (or zones) with that activity occurring, and/or the importance of the activity relative to other activities included in the analysis. Please see Appendix A for guidance on preparing and including information on intra- and inter-activity weights using qualitative (i.e., scores of 'more' or 'less' fishing in a cell, visitation or trip numbers for recreational activities) or quantitative (i.e., commercial fishing catch, effort level, revenues, profits) data.

Functionally, IS of pixel or management zone i is:

$$IS_i = \frac{1}{n}\sum_{i,j} U_{ij} I_j \tag{6.2}$$

where:

n = number of human use activities included in the analysis.

U_{ij} = usage or intra-activity weight (optional) of activity j in pixel or management zone i. If the user does not include intra-activity weights (i.e., model default), U_{ij} represents usage and is scored by presence (U_{ij} = 1) or absence (U_{ij} = 0) of the activity in the cell or zone. When intra-activity weights are included, U_{ij} reflects the weights as U_{ij} = X_{ij} / $Xmax_j$, where X_{ij} is the intra-activity weight of activity j in pixel or management zone i and $Xmax_j$ is the maximum intra-activity weight for all cells or zones where the activity occurs.

I_j = inter-activity weight (optional) of activity j relative to other activities included in the analysis. If the user treats all activities as equally important (model default), I_j is ignored (i.e., I_j = 1). When inter-activity weights are included, I_j reflects the weights as I_j = Y_j / *Ymax*, where Y_j is the inter-activity weight of activity j and $Ymax$ is the maximum inter-activity weight for all activities.

6.1.4 Limitations and Simplifications

This model is a very simple framework that provides little insight into how human activities might change under different scenarios of change in the coastal and marine environment. Such insights are best gleaned from models that include descriptors of human behavior. However, scenarios that add or remove activities or change weights of various activities and/or locations can be used to explore change.

> **Warning:** the model is very sensitive to inter- and intra-activity weights. Therefore, the assumptions you make when including these optional inter- and/or intra-activity weights will strongly affect model outputs. If you are unsure of how to appropriately include inter- or intra-activity weights, we encourage you to conduct several model runs to see how different weighting schemes affect model outputs.

6.1.5 Data Needs

The model uses an interface to input all required and optional model data. There are two options: the standard overlap analysis model that computes use intensity for each raster pixel, and an option to compute intesity by larger management zones. Each is a standalone model in InVEST, however the inputs required have the same descriptions and requirements so they are not reft below. Here we outline the options presented to the user via the interface, and the maps and data tables used by the model. First we describe required inputs, followed by a description of optional inputs.

Required Inputs

The required inputs are the minimum data needed to run this model. The minimum input data allows the model to run without importance weights or distance decay, both of which are optional parameters.

1. **Workspace Location (required)**. Users are required to specify a workspace folder path. We recommend that you create a new folder for each run of the model. For example, by creating a folder called "runBC" within the "OverlapAnalysis\Recreation" folder, the model will create "intermediate" and "output" folders within this "runBC" workspace. The "intermediate" folder will compartmentalize data from intermediate processes. The model's final outputs will be stored in the "output" folder.

```
Name: Path to a workspace folder.  Avoid spaces.
Sample path: \InVEST\OverlapAnalysis\BCrun
```

2. **Analysis Zones Layer (required).** A polygon shapefile that defines the area of interest for the standard analysis. The AOI must be projected with linear units equal to meters. For the management zones model, a similar shapefile is needed except the AOI should be divided into appropriate management zones.

```
Name: File can be named anything, but no spaces in the name
File type: Polygon shapefile (.shp)
Sample path: \InVEST\OverlapAnalysis\Input\AOI_WVCI.shp
```

3. **Analysis Cell Size (required).** This determines the spatial resolution at which the model runs and at which the results are summarized. For example, if you want to run the model and see results at a 100m x 100m pixel size then enter "100."

```
Name: A numeric text string (positive integer)
File type: text string (direct input)
Sample (default): 1000
```

4. **Overlap Analysis Data Directory (required).** Users are required to specify the path on their system to a folder containing only the input data for the Overlap Analysis model. Input data can be point, line or polygon data layers indicating where the human use activity takes place (e.g., whale watching, diving, or kayaking in a marine setting). Please note that optional intra-activity importance information, described below for optional input #1, can be associated with each layer. **In InVEST 3.1.0 and earlier, there may be no more than 32 layers in this directory.**

```
Name: Path to an activity data folder.  Avoid spaces.
Sample path: \InVEST\OverlapAnalysis\Input\RecreationLayers_RIS\
```

Note: All data in this folder must be shapefiles and projected in meters. For general help with creating and editing shapefiles, try documentation provided by ArcGIS or QGIS. For more specific InVEST-related GIS video tutorials, consider enrolling in the online course, Introduction to the Natural Capital Project Approach.

Optional Inputs

The next series of inputs are optional for added model functionality.

1. **Intra-Activity Attribute Name (optional).** The user has the option of providing information on the importance of locations (i.e., polygons or points) within a layer of human use data (e.g., one fishing ground may be much more valuable than another; certain kayaking routes may be more popular than others). These intra-activity importance scores can be qualitative or quantitative (see Appendix for further description of data inputs) and must be listed in a new column of the attribute tables for each layer included in the Overlap Analysis (see intra-activity weighting in *The Model* section). The name given to the column that contains the intra-activity importance scores must be the same for all layers contained within the directory specified by input #4. The model uses this information to weight the importance of areas found within each input layer.

```
Names: Text string containing letters and/or numbers (must start with a letter).
Field name must correspond to an existing column name in each layer's attribute table
Sample: RIS
```

2. **Inter-Activity Weight Table (optional).** The model also allows users to provide information on the relative importance of uses. This .csv file lists the activities and gives them a numerical relative importance weighting. The default files demonstrate the required structure; it is recommended that these files not be overwritten. In the .csv table, it is important that the name of each use exactly corresponds to the given name of the shapefile that represents that use.

```
Names: File can be named anything, but no spaces in the name
File type: Comma-separated values file (.csv)
Sample path: \InVEST\OverlapAnalysis\Input\Recreation_Inputs.csv
```

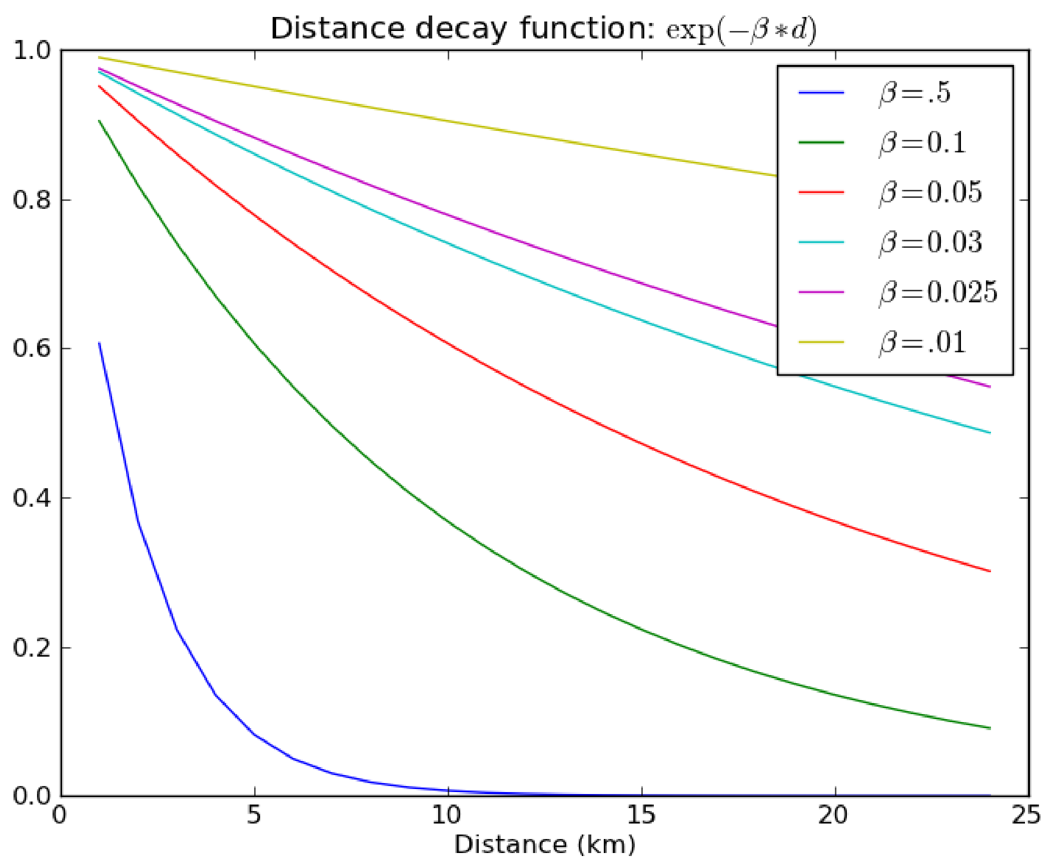
3. **Points Layer of Human Use Hubs (optional).** The model allows users to down-weight areas or zones used for different activities as a function of the distance from important land-based hubs such as ports, marinas, or public access points. This input GIS layer must be a point shapefile and projected in meters.

```
Names: File can be named anything, but no spaces in the name
File type: Point shapefile (.shp)
Sample path: \InVEST\OverlapAnalysis\Input\PopulatedPlaces_WCVI.shp
```

4. **Distance Decay Rate (optional).** If a GIS layer is specified for optional input #3, the model will use a decay rate of β =0.025 by default. If this input is not specified, no distance decay occurs and this rate is ignored. See Figure 1 for how changing this parameter changes the decay rate. With a decay rate of 0.025, an importance score of 1 would decrease to ~0.8 at a distance of approximately 10 km from the nearest hub. User judgment should be exercised when using this option. The following scenario illustrates one example of how users might use the distance decay function. Suppose you know that the intensity of human activities is greatest in areas relatively close to the ports, marinas, and other public access points, but you do not have the data necessary to construct spatially-explicit weighting factors to reflect this knowledge. In the absence of these data, the distance decay function could be used to reflect this intensity / distance tradeoff. You can choose a decay rate that reflects your best judgment on how the importance (e.g., intensity) of activities declines with distance from important population centers, marinas, or access points. For example, if most recreational fishing grounds are located within 10 km from the central marina, you could choose a decay parameter of β =0.01 to reflect a gradual threshold in the decline of importance of more distant sites, or β =0.5 to reflect a sharper threshold.

```
Names: A string of numeric text with a value between 0 and 1
File type: Text string (direct input to the ArcGIS interface)
Sample (default): 0.025
```

Exponential decay functions used to downweight importance of activities based on distance from land-based access point


Distance decay function: $\exp(-\beta * d)$
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
Distance (km)
$\beta = .5$
$\beta = 0.1$
$\beta = 0.05$
$\beta = 0.03$
$\beta = 0.025$
$\beta = .01$

Multiple Runs of the Model

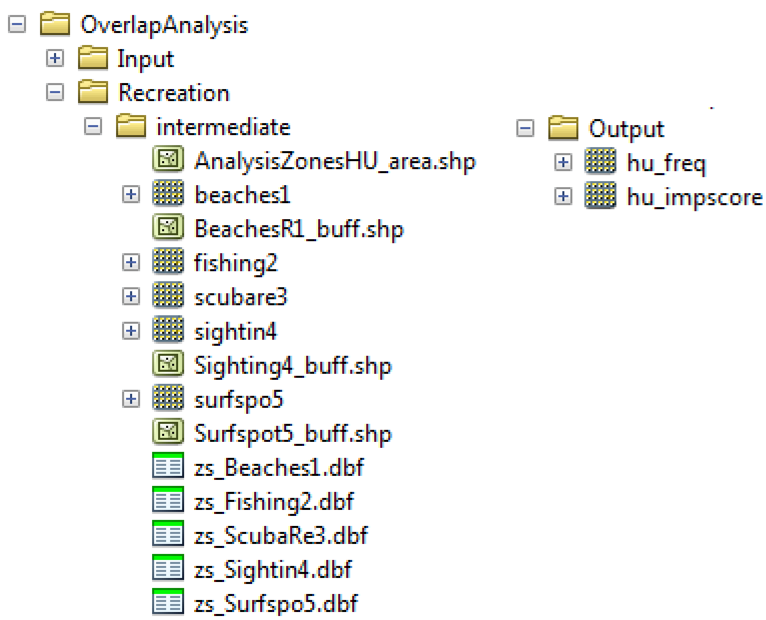
The tool setup is the same as for a single run, but you must specify a new workspace for each new run. Make sure each new workspace exists under the main workspace folder (i.e. *OverlapAnalysis* folder in the example above). As long as all data are contained within the main Input data folder you can use the same Input folder for multiple runs. For example, using the sample data, if you wanted to create two runs of the Overlap Analysis model based on two different weighting systems for fishing fleets, you could use the Input data folder under main Overlap Analysis folder and create two new workspace folders, runFisheries1 and runFisheries2.

6.1.6 Running the Model

The model is available as a standalone application accessible from the Windows start menu. For Windows 7 or earlier, this can be found under *All Programs -> InVEST +VERSION+ -> Overlap Analysis*. Windows 8 users can find the application by pressing the windows start key and typing "overlap" to refine the list of applications. The standalone can also be found directly in the InVEST install directory under the subdirectory *invest-3_x86/invest_overlap_analysis.exe*.

Viewing Output from the Model

Upon successful completion of the model, you will see new folders in your Workspace called "intermediate" and "Output". The Output folder, in particular, will contain several types of spatial data, which are described in the **Interpreting Results** section.



You can view the output spatial data in ArcMap using the Add Data button. 

You can change the symbology of a layer by right-clicking on the layer name in the table of contents, selecting "Properties", and then "Symbology". There are many options here to change the way the data appear in the map.

You can also view the attribute data of output files by right clicking on a layer and selecting "Open Attribute Table".

6.1.7 Interpreting Results

Model Outputs

The following is a short description of each of the outputs from the Overlap Analysis model. Each of these output files is saved in the "Output" folder that is saved within the user-specified workspace directory:

Output Folder

- Output\hu_freq
  - This raster layer depicts the frequency of activities for each cell or management zone for the study area. Each layer input is only counted once regardless of the number of features within that layer overlapping a cell. Therefore, if three layers are specified in the input directory, then the max value of this output is 3.
  - This is the default model output that will be generated for each run of the model.
- Output\hu_impscore
  - This raster layer depicts Importance Scores for each cell or management zone for the study area.
  - This output is only generated if the user includes intra-activity weights defined by optional input #1: "Importance Score Field Name".
- overlap_analysis-log-yr-mon-day-min-sec].txt
  - Each time the model is run a text file will appear in the workspace folder. The file will list log information that can be used to identify detailed configurations of each of scenario simulation.

6.1.8 Appendix A

Preparing Input Data

Maps of Fishing Grounds

Users should create a layer of polygons or points to define where individual fishing fleets operate. Fleets can be defined however you deem appropriate. Often, fleets are defined by their sector (e.g., commercial, recreational, subsistence), the species or species complexes they target (e.g., prawn, salmon, groundfish), and the gear that they use (e.g., trawl, seine, longline). For example, fleets might be commercial groundfish trawl, subsistence salmon seine, or recreational tuna hook and line.

For each fleet you decide to include, you must have information on where that fleet fishes. Locations can be points or polygons. You can generate these layers if existing maps of spatial distribution of fishing catch or effort are available to you. These maps are not often readily available, in which case, you can summarize catch, effort, or revenue data by management zone or statistical area. Availability of these data varies regionally – most regional management councils in the U.S. collect these data and make them publicly available through data clearinghouses associated with regional management councils (e.g., Pacific Fisheries Information Network associated with Pacific Fisheries Management Council). When summary by management zone or statistical area is unavailable, information can be solicited from stakeholders through exercises where they draw polygons or points on maps. If none of these are options for you, but you have habitat information available, it is possible to draw habitat-species-gear associations and coarsely estimate where fleet activity may occur.

Recreational Activity Layers

Spatially explicit data on recreation activities can be collected from a variety of sources including local tourism operators, government agencies, and guide books. In most areas, there is no clearinghouse for this type and users will likely need to combine data from a variety of sources.

Importance Data (Optional)

Intra-fleet Weights Quantitative or qualitative or data on which locations in the coastal and marine environment are most or least important for a human use (i.e., intra-activity weights) can be easily prepared and included in the Overlap Analysis model. Whichever type of data is used does not need to be consistent across human use activities. For example, when spatially-explicit catch data exist for one fishing fleet, and another fishing fleet only has qualitative rankings of importance of different fishing grounds, both data sets can be used. Intra-fleet weights are entered for each polygon or point in each data layer's attribute table. If intra-fleet weights are missing for one or more data layers in the analysis, users must include a placeholder column (i.e., values for all polygons in the layer = 1) for the model to run correctly.

Quantitative data are likely to be catch, effort, profit, or revenue information for fisheries. For recreation, the number of trips or number of visitors to each site is the suggested metric to be used to weight activities. Alternatively, users may use the number of days that an area is open to particular activities or other metrics that proxy for importance or usage. Higher values should indicate polygons or points of higher importance than those with lower values.

Qualitative scoring is a good option for users without quantitative input data. Low scores should indicate least important locations for the activity, high scores most important areas, and multiple areas should be allowed to have the same score (i.e., areas are given scores, not ranks). We encourage users to take care in assignment of values to locations as these values strongly influence outcomes. For example, if one fishing area polygon is given a score of 1, and another a score of 2, is the 2nd polygon twice as "important" as the first? If not, and the two polygons are more similar in their importance, the user could considering scoring more closely to one another (e.g., score of 1.75 and 2, instead of 1 and 2) or score on a larger scale (e.g., scores of 4 and 5, instead of 1 and 2). The onus is on the user to decide which range of weights to use. If you are unsure of how to appropriately include these weights, we encourage you to conduct several model runs to see how different weighting schemes affect model outputs. A common method for obtaining qualitative information on the importance of an activity is by querying stakeholders or decision-makers in the region. InVEST will soon include a mapping tool to help collect data from stakeholders. The tool will include functionality for entering intra-activity weights. If using the InVEST drawing tool (forthcoming) while querying stakeholders, importance scores can be input when generating layers.

Once intra-activity weights are input into the model, they are scaled by the maximum value for all locations where the activity occurs. For example, if the user has identified 3 fishing grounds for a fleet, with values of 2, 4, and 5, they will be scaled by 5, to be 0.4, 0.8 and 1.0.

Inter-activity Weights The user has the option to include information on the importance of activities relative to one another so that all activities are not treated equally. This information is not spatially explicit, rather is in the form of one value for each activity. If the user chooses to include inter-activity weights, they must be included for all activities. Inter-activity weights can be qualitative (e.g., stakeholder designated) or quantitative (e.g., total catch, effort, profit, or revenue; socio-economic assessment of contributions of each fishing fleet to community stability or tax base), but the same metric should be used to weight all activities. For recreation, if the user does not have spatially explicit data on numbers of recreation trips, but does have the aggregate number of trips or participants for each activity, these numbers can be used to construct an importance ranking of each activity by using the percentage of trips / participants in each activity as inter-activity weights. For fisheries, for example, if running the model for three fishing fleets, inter-activity weights could be calculated using total revenue earned by each fleet as is done in the example presented earlier in this chapter. It would be inappropriate to determine weights by comparing one fleet's catches to the others' revenues. Given this caution, when determining inter-activity weights, users should choose a common quantitative (e.g., catch, revenue for fishing fleets) or qualitative (e.g., scores from stakeholder input) metric that is applicable across all activities. Similar to the intra-activity weights, inter-activity weights are not ranks (i.e., activities can have the same weights), and must be included for all data layers. Once input into the model, quantitative or qualitative values are scaled by the maximum value for all activities.

The caution in the preceding, intra-activity, section about the numeric scales used for qualitatively weighting activities applies here, as weights strongly affect model outputs. To reiterate, using a hypothetical model run for recreational data, if the inter-activity weight for whale-watching is 1, and kayaking 2, is the kayaking twice as "important" as whale-watching? If the activities are actually more similar, the weights should be closer to one another (e.g., score of 1.75 and 2, instead of 1 and 2) or score on a larger scale (e.g., scores of 4 and 5, instead of 1 and 2). Users are responsible for choosing the range of weights to use, and we encourage you to conduct several model runs to see how different weighting schemes affect model outputs.

6.1.9 References

Beck, M.W, Z. Ferdana, J. Kachmar, K. K. Morrison, P. Taylor and others. 2009. Best Practices for Marine Spatial Planning. The Nature Conservancy, Arlington, VA. 32 pp.

CDFG (California Department of Fish and Game). 2008. California Marine Life Protection Act. Master Plan for Marine Protected Areas. 110 pp.

CRMC (Coastal Resources Management Council). 2010. Rhode Island Ocean Special Area Management Plan: Adopted by the Rhode Island Coastal Resources Management Council October 2010. 993 pp.

DFO (Department of Fisheries and Oceans). 2008. Canadian Fisheries Statistics 2006. Ottawa: Fisheries and Oceans Canada.

GBRMPA (Great Barrier Reef Marine Park Authority). 2003. Great Barrier Reef Marine Park Zoning Plan 2003. Australian Government. 220 pp.

6.2 Coastal Vulnerability Model

6.2.1 Summary

Faced with an intensification of human activities and a changing climate, coastal communities need to better understand how modifications of the biological and physical environment (i.e. direct and indirect removal of natural habitats for coastal development) can affect their exposure to storm-induced erosion and flooding (inundation). The InVEST Coastal Vulnerability model produces a qualitative estimate of such exposure in terms of a Vulnerability Index, which differentiates areas with relatively high or low exposure to erosion and inundation during storms. By coupling these results with global population information, the model can show areas along a given coastline where humans are most vulnerable to storm waves and surge. The model does not take into account coastal processes that are unique to a region, nor does it predict long- or short-term changes in shoreline position or configuration.

Model inputs, which serve as proxies for various complex shoreline processes that influence exposure to erosion and inundation, include: a polyline with attributes about local coastal geomorphology along the shoreline, polygons representing the location of natural habitats (e.g., seagrass, kelp, wetlands, etc.), rates of (observed) net sea-level change, a depth contour that can be used as an indicator for surge level (the default contour is the edge of the continental shelf), a digital elevation model (DEM) representing the topography and (optionally) the bathymetry of the coastal area, a point shapefile containing values of observed storm wind speed and wave power, and a raster representing population distribution.

Outputs can be used to better understand the relative contributions of these different model variables to coastal exposure and highlight the protective services offered by natural habitats to coastal populations. This information can help coastal managers, planners, landowners and other stakeholders identify regions of greater risk to coastal hazards, which can in turn better inform development strategies and permitting. As the results provide a qualitative representation of erosion and inundation risks rather than quantifying shoreline retreat or inundation limits.

6.2.2 Introduction

Coastal regions, which constantly subject to the action of ocean waves and storms, naturally experience erosion and inundation over various temporal and spatial scales. However, coastal erosion and inundation pose a threat to human populations, activities and infrastructure, especially within the context of a changing climate and increasing coastal populations. Moreover, these increases in anthropogenic pressure can lead to the loss and degradation of coastal ecosystems and their ability to provide protection for humans during storms. Thus, it is important to understand the role of various biological and geophysical factors in increasing or decreasing the threat of coastal erosion and inundation in order to better plan for future development. In particular, it is important to know how natural habitats can mitigate the forces responsible for coastal erosion and inundation so that management actions might best preserve the protective services provided by coastal ecosystems.

A number of models estimate the vulnerability of coastal regions to long-term sea level rise, erosion and inundation based on geophysical characteristics (Gornitz et al. 1991, Hammar-Klose and Thieler 2001, Cooper and McLaughlin 1998). There are also methods to qualitatively estimate the relative role natural habitats play in reducing the risk of erosion and inundation of particular areas (WRI 2009, Bush et al. 2001). However, few models map the relative vulnerability of coastal areas to erosion and inundation based on both the geophysical and natural habitat characteristics of a region. It is our aim to fill that gap with the Coastal Vulnerability model.

The Coastal Vulnerability model produces a qualitative index of coastal exposure to erosion and inundation as well as a map of the location and size of human settlements. The model does not value directly any ecoystem service, but ranks sites as having a relatively low, moderate or high risk of erosion and inundation. It is relatively simple to use and quick to run, and it can be applied in most regions of the world with data that are, for the most part, relatively easy to obtain.

Model outputs are easy to understand and the spatial coverage of the results allows these outputs to be overlaid with other spatial information for users to perform further analysis as they see fit. By highlighting the relative role of natural habitat at reducing exposure and showing the areas where coastal populations are threatened, the model can be used, in a simple way, to investigate how some management action or land use change can affect the exposure of human populations to erosion and inundation.

6.2.3 The Model

The InVEST Coastal Vulnerability model produces an exposure index raster and a coastal population raster. The exposure index raster contains ranks of the relative exposure of different coastlines segments to erosion and inundation caused by storms within the same coastal region of interest. The coastal population raster shows the distribution of human population density within the coastal region of interest. These maps are constructed using seven bio-geophysical variables and a population raster. The seven bio-geophysical variables represent the natural biological and geomorphic characteristics of a region, the amount of expected net sea-level rise and the relative wind and wave forcing associated with storms. As the model assesses the relative exposure of different areas within the domain of interest, model outputs are relevant when computed for a relatively large and/or non-uniform coastal region. Coupled together, the exposure index and coastal population rasters can be used to create maps that show the relative vulnerability of human populations to coastal storms.

How it Works

The model creates the exposure index and coastal population maps using a spatial representation (raster) of population and spatial representations (shapefiles and rasters) of seven bio-geophysical variables:

1. Geomorphology
2. Relief
3. Natural habitats (biotic and abiotic)
4. Net sea level change

5. Wind Exposure
6. Wave Exposure
7. Surge potential depth contour

The outputs of the model are rasters that overlay the shoreline in the coastal region of interest, with a spatial resolution defined by the user ($\geq$ 250 meters). These rasters contain a number of indices and rankings of input variables (described below) and can be used to the create maps that fit the user's needs. Below are details describing the model variables and how the output rasters are created.

Exposure Index

The model computes the physical exposure index by combining the ranks of the seven biological and physical variables at each shoreline segment. Ranks vary from very low exposure (rank=1) to very high exposure (rank=5), based on a mixture of user- and model-defined criteria (see *Table 4.1*). This ranking system is based on methods proposed by Gornitz et al. (1990) and Hammar-Klose and Thieler (2001).

Table 4.1

Rank	Very Low	Low	Moderate	High	Very High
Variable	1	2	3	4	5
Geomorphology	Rocky; high cliffs; fjord; fiard, seawalls	Medium cliff; indented coast, bulkheads and small seawalls	Low cliff; glacial drift; alluvial plain, revetments, rip-rap walls	Cobble beach; estuary; lagoon; bluff	Barrier beach; sand beach; mud flat; delta
Relief	0 to 20 Percentile	21 to 40 Percentile	41 to 60 Percentile	61 to 80 Percentile	81 to 100 Percentile
Natural Habitats	Coral reef; mangrove; coastal forest	High dune; marsh	Low dune	Seagrass; kelp	No habitat
Sea Level Change	0 to 20 Percentile	21 to 40 Percentile	41 to 60 Percentile	61 to 80 Percentile	81 to 100 Percentile
Wave Exposure	0 to 20 Percentile	21 to 40 Percentile	41 to 60 Percentile	61 to 80 Percentile	81 to 100 Percentile
Surge Potential	0 to 20 Percentile	21 to 40 Percentile	41 to 60 Percentile	61 to 80 Percentile	81 to 100 Percentile

Table 4.1: List of Bio-Geophysical Variables and Ranking System for Coastal Exposure.

The model calculates the exposure index EI for each shoreline segment as the geometric mean of all the variable ranks:

$$EI = (R_{Geomorphology}R_{Relief}R_{Habitats}R_{SLR}R_{WindExposure}R_{WaveExposure}R_{Surge})^{1/7} \tag{6.3}$$

or more generally:

$$EI = \left(\prod_{i=1}^{n} R_i\right)^{1/n} \tag{6.4}$$

where R_i represents the ranking of the i^{th} bio-geophysical variable that is provided by the user to calculate EI.

In addition to mapping the exposure index, the model computes an erosion index ErI as:

$$ErI = (R_{Geomorphology} R_{Habitats} R_{WaveExposure})^{1/3} \tag{6.5}$$

We designed this additional output to invite users to explore the different ways in which model results vary with different combinations of variables. Here, we have assumed that the most important factors affecting erosion are geomorphology, natural habitats and wave height/period. The erosion index provides an example for how users can similarly create their own index by defining a unique combination of these rankings. In the remainder of this section, we will provide a more detailed description of the variables presented in *Table 4.1*.

Additionally, we provide all intermediate and raw results computed by the model so users can generate outputs using different ranking or computation methods.

Geomorphology

Rocky cliffs are less prone to erosion and inundation than bluffs, beaches or deltas. Consequently, a relative ranking of exposure scheme based on geomorphology similar to the one proposed by Hammar-Klose and Thieler (2001) has been adopted. Supplied in *Appendix A* is a definition of the terms used in this classification, which applies mostly to the North American continent. This classification will be expanded to cover more regions of the world in later versions of this model.

Note that we included structures in this list of features because they are present along most developed coasts.

If the user's geomorphology source has more categories than the ones presented in *Table 4.1*, it is left to the users discretion to reclassify their data to match the provided ranking system, as explained in the *Data Needs* section, and in *Appendix B*. It is recommend however, that the user include shore parallel hard structures (seawalls, bulkheads, etc) in this classification and that they apply a low to moderate rank (1-3), depending on their characteristics. For example, a large, concrete seawall should be assigned a rank 1 as they are typically designed to prevent inundation during storm events and are designed to withstand damage or failure during the most powerful storms. It is recommended that low revetments or riprap walls be assigned a rank of 3 as they do not prevent inundation and may fail during extreme events.

The ranking presented in the above table is but a suggestion. Users can change the ranking of different shoreline types as they see fit by following directions presented in the *Data Needs* section.

The model requires a polyline shapefile that runs along the coastline of interest. This shapefile must be discretized into different segments where the geomorphology rank changes. The attribute table of this shapefile requires a field called RANK of type 'Short Integer', which contains the numeric rank (1-5) of the geomorphology type along each particular segment. As mentioned above and explained later in this document, the value of the RANK associated with a particular shoreline type is left to the discretion of the user.

Relief

Sites that are, on average, at greater elevations above Mean Seal Level (MSL) are at a lower risk of being inundated than areas at lower elevations. Relief is defined in the model as the average elevation of the coastal land area that is within a user-defined radius (default = 5 km) from each shore segment of the discretized shoreline. This resolution was chosen because of the relative coarseness of most freely available terrestrial digital elevation models (DEMs).

For this variable, the model requires a DEM that covers the area of interest. As a part of the InVEST download package, a global elevation DEM is provided. The accuracy and quality of this dataset varies. If users have access to a more local, accurate data source, we encourage them to use it rather than the provided DEM.

Natural Habitats

Natural habitats (marshes, seagrass beds, mangroves, coastal dunes) play a vital role in decreasing the impacts of coastal hazards that can erode shorelines and harm coastal communities. For example, large waves break on coral reefs before reaching the shoreline, mangroves and coastal forests dramatically reduce wave heights in shallow waters, and decrease the strength of wave- and wind-generated currents, seagrass beds and marshes stabilize sediments and encourage the accretion of nearshore beds as well as dissipate wave energy. On the other hand, beaches with little to no biological habitats or sand dunes offer little protection to erosion and inundation.

The ranking proposed in *Table 4.1* is based on the fact that fixed and stiff habitats that penetrate the water column (e.g., coral reefs, mangroves) and sand dunes are the most effective in protecting coastal communities. Flexible and seasonal habitats, such as seagrass, reduce flows when they can withstand their force, and encourage accretion of sediments. Therefore, these habitats receive a lower ranking than fixed habitats. It is left to the user's discretion to separate sand dunes into high and low categories. It is suggested, however, that since category 4 hurricanes can create a 5m surge height, 5m is an appropriate cut-off value to separate high (>5m) and low (<5m) dunes. If the user has local knowledge about which habitats and dune elevations provide the better protection in their area of interest, they are free to deviate from these recommendations for their application.

It is important to note that this ranking is based on the result of extensive literature review and the author's best judgment. Users are free to modify it as they see fit.

To compute a Natural Habitat exposure rank for a given shoreline segment, the model determines whether a certain class of natural habitat (*Table 4.1*) is within a user-defined search radius from the segment. (See Section 2 and *Appendix B* for a description of how the model processes natural habitat input layers.) When all N habitats fronting that segment have been identified, the model creates a vector *R* that contains all the ranks $R_k, 1 \leq k \leq N$, associated with these habitats, as defined in *Table 4.1*. Using those rank values, the model computes a final *Natural Habitat* exposure rank for that segment with the following formulation:

$$R_{Hab} = 4.8 - 0.5\sqrt{(1.5 \max_{k=1}^{N}(5 - R_k))^2 + (\sum_{k=1}^{N}(5 - R_k)^2 - \max_{k=1}^{N}(5 - R_k))^2)} \tag{6.6}$$

where the habitat that has the lowest rank is weighed 1.5 times higher than all other habitats that are present near a segment. The final ranking values vary between a maximum of 4 when a segment is solely fronted by kelp or seagrass, to a minimum of 1.025 when it is fronted by a mangrove and coastal forests, a seagrass bed and a coral reef. This formulation allows us to maximize the accounting of the protection services provided by all natural habitats that front a shoreline segment. In other words, it ensures that segments that are fronted or have only one type of habitat (e.g., high sand dune) are more exposed than segments with more than one habitat (e.g., coral reefs and high sand dune). See *Appendix B* for a detailed account of all possible final rank values that can be obtained with equation (6.6).

To include this variable in the exposure index calculation, the model requires separate polygon shapefiles representing each natural habitat type within the area of interest, along with a csv file that contains the name, rank and search radius of each of the shapefile. As mentioned above, we left it to the users' discretion to modify the ranks of the natural habitat layers as they see fit. We present a complete description of the requirements for this variable and instructions on how to prepare this variable for the model in *Appendix B*.

Net Sea-Level Change

The relative net sea level change along the coastline of a given region is the sum of global sea level rise (SLR), local SLR (eustatic rise) and local land motion (isostatic rise). This results in net sea level change values that can be positive (sea level is rising) or negative (sea level is decreasing) at a particular site. To include this variable in the exposure index calculation, the model takes either a polygon shapefile where polygons delineate the extents of a uniform sea level change, or a point shapefile where the points carry the recorded sea level change. In either case, the model will look for a field named 'Trend', which is the yearly rate of sea level change (usually given in mm/yr). Please consult *Appendix B* for suggestions of how to create this input.

Wave Exposure

The relative exposure of a reach of coastline to storm waves is a qualitative indicator of the potential for shoreline erosion. A given stretch of shoreline is generally exposed to either oceanic or locally-generated wind-waves. Also, for a given wave height, waves that have a longer period have more power than shorter waves. Coasts that are exposed to the open ocean generally experience a higher exposure to waves than sheltered regions because winds blowing over a very large distance, or fetch, generate larger waves. Additionally, exposed regions experience the effects of long period waves, or swells, that were generated by distant storms.

The model estimates the relative exposure of a shoreline segment to waves E_w by assigning it the maximum of the weighted average power of oceanic waves, $E_w^o`andlocallywind-generatedwaves, :math:`E_w^l$:

$$E_w = \max(E_w^o, E_w^l) \tag{6.7}$$

For oceanic waves, the weighted average power is computed as:

$$E_w^o = \sum_{k=1}^{16} H[F_k] P_k^o O_k^o \tag{6.8}$$

where $H[F_k]$ is a heaviside step function for all of the 16 wind equiangular sectors *k*. It is zero if the fetch in that direction is less than 60 km, and 1 if the fetch is equal to 60 km (for computational reason, we compute fetch distances up to 60 km):

$$H[F_k] = \begin{cases} 0 & \text{if } F_k < 60km \\ 1 & \text{if } F_k = 60km \end{cases} \tag{6.9}$$

In other words, this function only considers angular sectors where oceanic waves (assuming sheltered water bodies have fetch lengths less than 50km) have the potential to reach the shoreline in the evaluation of oceanic wave exposure. Further, $P_k^o O_k^o$ is the average of the highest 10% wave power values (P_k^o) that were observed in the direction of the angular sector *k*, weighted by the percentage of time (O_k^o) when those waves were observed in that sector. For all waves in each angular sector, wave power is computed as:

$$P = \frac{1}{2} H^2 T \tag{6.10}$$

where $P[kW/m]$ is the wave power of an observed wave with a height $H[m]$ and a period $T[s]$.

For locally wind-generated waves, E_w^l is computed as:

$$E_w^l = \sum_{k=1}^{16} P_k^l O_k^l \tag{6.11}$$

which is the sum over the 16 wind sectors of the wave power generated by the average of the highest 10% wind speed values P_k^l that propagate in the direction *k*, weighted by the percent occurrence O_k^l of these strong wind in that sector.

The power of locally wind-generated waves is estimated with Equation (6.10). The wave height and period of the locally generated wind-waves are computed for each of the 16 equiangular sectors as:

$$\begin{cases} H = \widetilde{H}_\infty \left[\tanh\left(0.343 \widetilde{d}^{1.14}\right) \tanh\left(\frac{2.14 \cdot 10^{-4} \widetilde{F}^{0.79}}{\tanh(0.343 \widetilde{d}^{1.14})} \right) \right]^{0.572} \\ \\ T = \widetilde{T}_\infty \left[\tanh\left(0.1 \widetilde{d}^{2.01}\right) \tanh\left(\frac{2.77 \cdot 10^{-7} \widetilde{F}^{1.45}}{\tanh(0.1 \widetilde{d}^{2.01})} \right) \right]^{0.187} \end{cases} \tag{6.12}$$

where the non-dimensional wave height and period $\widetilde{H}_\infty$ and $\widetilde{T}_\infty$ are a function of the average of the highest 10% wind speed values $U[m/s]$ that were observed in in a particular sector: $\widetilde{H}_\infty = 0.24U^2/g$, and $\widetilde{T}_\infty = 7.69U^2/g$, and where the non-dimensional fetch and depth, $\widetilde{F}_\infty$ and $\widetilde{d}_\infty$, are a function of the fetch distance in that sector $F[m]$ and the average water depth in the region of interest $d[m]$: $\widetilde{F}_\infty = gF/U^2$, and $\widetilde{T}_\infty = gd/U^2$. $g[m/s^2]$ is the acceleration of gravity.

This expression of wave height and period assumes fetch-limited conditions, as the duration over which the wind speed,:math:U, blows steadily in the direction of the fetch, F (USACE, 2002; Part II Chap 2). Hence, model results might over-estimate wind-generated waves characteristics at a site.

Since sheltered areas of the coast (areas that are within embayments or sheltered from oceanic waves by geomorphic features) are not exposed to oceanic waves ($E_w^o = 0$) the relative exposure to waves is simply $E_w = E_w^l$. In order to differentiate between exposed and sheltered areas , the model uses a fetch filter; segments for which two or more of the 16 fetches do not exceed a user-defined threshold distance are assumed to be sheltered.

As a part of the InVEST download package, a shapefile with default wind and wave data compiled from 8 years of WAVEWATCH III (WW3, Tolman (2009)) model hindcast reanalysis results is provided. As discussed in the previous section, for each of the 16 equiangular wind sector, the average of the highest 10% wind speed, wave height and wave power have been computed. If users wish to use another data source, we recommend that they use the same statistics of wind and wave (average of the highest 10% for wind speed, wave height and wave power), but they can use other statistics as well. However, these data must be contained in a point shapefile with the same attribute table as the WW3 data provided.

The model differentiates from exposed and sheltered areas by using a combination of user-defined fetch distance threshold and depth threshold. If the number of fetch vectors or average depth over fetch vectors is greater than the user-input exposure proportion value, the coastal segment associated with those fetch and depth values will be deemed sheltered (see *Data Needs* section). We provide an example of how to estimate the fetch threshold distance in *Appendix B*. The depth threshold distance can be estimated using the "information" tool in GIS and scanning average depths in shallow, sheltered estuaries. Note that the distinction between sheltered and exposed thus obtained is qualitative.

Surge Potential

Storm surge elevation is a function of wind speed and direction, but also of the amount of time wind blows over relatively shallow areas. In general, the longer the distance between the coastline and the edge of the continental shelf at a given area during a given storm, the higher the storm surge. Unless a user decides to specify a certain depth contour appropriate to their region of interest, the model estimates the relative exposure to storm surges by computing the length of the continental shelf fronting an area of interest (otherwise, it computes the distance between the shoreline and the user-specified contour). For hurricanes in the Gulf of Mexico, a better approximation might be made by considering the distance between the coastline and the 30 meters depth contour (Irish and Resio 2010).

The model assigns a distance to all segments within the area of interest, even to segments that seem sheltered because they are too far inland, protected by a significant land mass, or on a side of an island that is not exposed to the open ocean.

Wind Exposure

Strong winds can generate high surges and/or powerful waves if they blow over an area for a sufficiently long period of time. The wind exposure variable is an optional output that ranks shoreline segments based on their relative exposure to strong winds. Wind exposure results are located in the *Intermediate* output folder. We compute this ranking by computing and mapping the Relative Exposure Index (REI; Keddy, 1982). This index is computed by taking the time series of the highest 10% wind speeds from a long record of measured wind speeds, dividing the compass rose (or the 360 degrees compass) into 16 equiangular sectors and combining the wind and fetch (distance over which wind blows over water) characteristics in these sectors as:

$$REI = \sum_{n=1}^{16} U_n P_n F_n \tag{6.13}$$

where:

- U_n is the average wind speed, in meters per second, of the 10% wind speeds in the n^{th} equiangular sector
- P_n is the percent of all wind speeds in the record of interest that blow in the direction of the n^{th} sector
- F_n is the fetch distance, in meters, in the n^{th} sector

For a given coastline segment, the model estimates fetch F distances over each of the 16 equiangular sectors by taking the average of k fetch segments within a section following (Keddy, 1982):

$$F = \frac{\sum_{n=1}^{k} f_n \cos\theta}{\sum_{n=1}^{k} \cos\theta} \tag{6.14}$$

where f_n is the n^{th} radial distance in an equiangular sector, and $\theta = 22.5deg/k$ (we recommend that $k \leq 9$).

Note that, in this model, wind direction is the direction winds are blowing FROM, and not TOWARDS. If users provide their own data, they must ensure that the data matches this convention before applying those data to this model. Also, note that, for computational reason, we compute fetch distances up to 60 km.

Social Exposure

When estimating the exposure of coastlines to erosion and inundation due to storms, it is important to consider the population of humans that will be subject to those coastal hazards. The Coastal Vulnerability model extracts population values along the shoreline at discrete segments with the user-specified spatial resolution. To obtain this raster showing the estimated number of people residing on a coastal area, the model overlays a raster containing population values pulled from a user-defined radius (see coastal neighborhood) at each grid cell with the rasterized shoreline. The model then assigns each discrete shoreline segment a population value by extracting the population value from the grid cell that overlaps the shoreline segment. As a part of the InVEST download package, a global population raster is provided with population values obtained from country level census data. As is the case with all input data, the user may provide their own population raster (e.g., *LandScan data <http://web.ornl.gov/sci/landscan/landscan_data_avail.shtml>*) if they have more accurate, local information.

6.2.4 Limitations and Simplifications

Beyond technical limitations, the exposure index also has theoretical limitations. One of the main limitations is that the dynamic interactions of complex coastal processes occurring in a region are overly simplified into the geometric mean of seven variables and exposure categories. We do not model storm surge or wave field in nearshore regions. More importantly, the model does not take into account the amount and quality of habitats, and it does not quantify the role of habitats are reducing coastal hazards. Also, the model does not consider any hydrodynamic or sediment transport processes: it has been assumed that regions that belong to the same broad geomorphic exposure class behave in a similar way. Additionally, the scoring of exposure is the same everywhere in the region of interest; the model does not take into account any interactions between the different variables in *Table 4.1*. For example, the relative exposure to waves and wind will have the same weight whether the site under consideration is a sand beach or a rocky cliff. Also, when the final exposure index is computed, the effect of biogenic habitats fronting regions that have a low geomorphic ranking are still taken into account. In other words, we assume that natural habitats provide protection to regions that are protected against erosion independent of their geomorphology classification (i.e. rocky cliffs). This limitation artificially deflates the relative vulnerability of these regions, and inflates the relative vulnerability of regions that have a high geomorphic index.

The other type of model limitations is associated with the computation of the wind and wave exposure. Because our intent is to provide default data for users in most regions of the world, we had to simplify the type of input required to compute wind and wave exposure. For example, we computed storm wind speeds in the WW3 wind database that we provide by taking the average of winds speeds above the 90th percentile value, instead of using the full time series of wind speeds. Thus we do not represent fully the impacts of extreme events. Also, we estimate the exposure to oceanic waves by assigning to a coastal segment the waves statistics of the closest WW3 grid point. This approach neglects any 2D processes that might take place in nearshore regions and that might change the exposure of a region. Similarly, we compute exposure in sheltered region by combining the average depth near a particular segment to the wind speed and direction in a sector, instead of modeling the growth and evolution of wind waves near that segment.

Consequently, model outputs cannot be used to quantify the exposure to erosion and inundation of a specific coastal location; the model produces qualitative outputs and is designed to be used at a relatively large scale. More importantly, the model does not predict the response of a region to specific storms or wave field and does not take into account any large-scale sediment transport pathways that may exist in a region of interest.

6.2.5 Data Needs

The model uses an interface to input all required and optional data, as outlined in this section. It outputs a link to an HTML file where the user can upload their results into a coastal vulnerability dashboard (http://vulpes.sefs.uw.edu/ttapp/cv-dash.php). The user can click through results on an interactive map and view surge potential, wave exposure, coastal exposure and more. To compute the exposure index the user has the option of uploading any or all of the variables in *Table 4.1*, with the exception of the wind-wave input layer and the bathymetry: *the model will not run unless a wind-wave input layer and DEM have been uploaded*.

Below, we outline the options that we offer to users in the interface, and the content and format of the required and optional input data that the model uses. we provide more information on how to fill the input interface, or on how to obtain data in *Appendix B*.

1. **Output area.** Specify whether all or only the sheltered shoreline segments appear in the output. This option has no effect on the computation performed by the model, and only affects the shore segments that appear in the output files.

2. **Workspace Location (required).** The user is required to specify a workspace directory path. It is recommended to create a new directory for each run of the model. The model will create an “intermediate” and an “output” directory within this workspace. The “intermediate” directory will compartmentalize data from intermediate processes. The model’s final outputs will be stored in the “output” directory.

```
Name: Path to a workspace directory. Avoid spaces.
Sample path: \InVEST\coastal_vulnerability
```

3. **Area of Interest (AOI, required).** Users must create a polygon feature layer that defines the Area of Interest (AOI). An AOI instructs the model where to clip the Land Polygon input data (inputs #2-3) in order to define the spatial extent of the analysis. The model uses the AOI’s projection to set the projection for the sequential intermediate and output data layers. We also recommend that they have a WGS84 datum. In order to allocate wind and wave information from the Wave Watch 3 data (WW3), this AOI must also overlap one or more of the provided WW3 points. If users are including the Surge Potential variable in the computation of the exposure index, the depth contour specified in the Coastal Vulnerability model must be specified, and the AOI must intersect that contour. If the AOI does not intersect that contour, the model will stop and provide feedback.

```
Name: File can be named anything, but no spaces in the name
File type: polygon shapefile (.shp)
Sample path: \InVEST\CoastalProtection\Input\AOI_BarkClay.shp
```

4. **Land Polygon (required).** This input provides the model with a geographic shape of the coastal area of interest, and instructs it as to the boundaries of the land and seascape. A global land mass polygon shapefile is provided as default (Wessel and Smith, 1996), but other layers can be substituted. If users have a more accurate, local polygon shapefile representing land masses, they are encouraged to use this data rather than the provided shapefile.

```
Name: File can be named anything, but no spaces in the name
File type: polygon shapefile (.shp)
Sample path (default): \InVEST\Base_Data\Marine\Land\global_polygon.shp
```

5. **Bathymetry layer. (required)** This input is used to compute the average depth along the fetch rays to determine the exposure of each shoreline segment (*Table 4.1*), and in the computation of surge potential. It should consist of depth information of bodies of water within the AOI as marked by the land polygon shapefile.

```
Name: File can be named anything, but no spaces in the name
File type: raster dataset
Sample path: \InVEST\Base_Data\Marine\DEMs\claybark_dem
```

6. **Layer value if path omitted (optional).** Integer value between 1 and 5. If bathymetry is omitted, replace all shore points for this layer with a constant rank value in the computation of the coastal vulnerability index. If both the file and value for the layer are omitted, the layer is skipped altogether.

```
Name: A positive integer between 1 and 5.
File type: text string (direct input to the interface)
Sample (default): empty
```

7. **Relief (required).** Digital Elevation Model (DEM). This input is used to compute the Relief ranking of each shoreline segment (*Table 4.1*). It should consist of elevation information covering the entire land polygon within the AOI. Focal statistics are computed on the input DEM within a range defined by the user (see Elevation averaging radius). The average of elevation values within this range is ranked relative to all other coastline segments within the AOI. Although the default raster for this layer is the same as for Bathymetry, each entry can refer to a separate raster, where one computes elevations above water, and the other below water.

```
Name: File can be named anything, but no spaces in the name
File type: raster dataset
Sample path: \InVEST\Base_Data\Marine\DEMs\claybark_dem\hdr.adf
```

8. **Layer value if path omitted (optional).** Integer value between 1 and 5. If relief is omitted, replace all shore points for this layer with a constant rank value in the computation of the coastal vulnerability index. If both the file and value for the layer are omitted, the layer is skipped altogether.

```
Name: A positive integer between 1 and 5.
File type: text string (direct input to the interface)
Sample (default): empty
```

9. **Elevation averaging radius (meters, required).** This input determines the radius around within which to compute the average elevation for relief.

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the interface)
Sample (default): 5000
```

10. **Mean sea level datum (meters, required).** This input is the elevation of Mean Sea Level (MSL) datum relative to the datum of the bathymetry layer that they provide. The model transforms all depths to MSL datum by subtracting the value provided by the user to the bathymetry. This input can be used to run the model for a future sea-level rise scenario.:

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the interface)
Sample (default): 0
```

11. **Model resolution (segment size in meters, required).** This input determines the spatial resolution at which the model runs and the resolution of the output rasters. To run the model at 250 x 250 meters grid cell scale, users should enter "250". A larger grid cell will yield a lower resolution, but a faster computation time (computation is in the order of :math:'O(n^3)' with n being the number of rows or columns in the raster).

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the interface)
Sample (default): 250
```

12. **Rays per sector (required).** Number of rays used to sample the ocean depth and land proximity within each of the 16 equiangular fetch sectors.

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the interface)
Sample (default): 1 (maximum = 9)
```

13. **Fetch Distance Threshold (meters, required).** Used to determine if the current segment is enclosed by land. This input is used in conjunction with the average ocean depth and exposure proportion to differentiate sheltered and exposed shoreline segments.:

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the interface)
Sample (default): 60000
```

14. **Depth Threshold (meters, required).** Used to determine if the current segment is surrounded by deep water. This input is used in conjunction with the fetch distance threshold and exposed segment to differentiate between sheltered and exposed shoreline segments.

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the interface)
Sample (default): 0
```

15. **Exposure proportion (meters, required).** The model uses this input (between 0.0 and 1.0) to determine if shore segments are exposed or sheltered. This is done in four steps:

    (a) Compute the number of fetch rays (N) that correspond to the proportion N: :math:'segments over water * exposure proportion'

    (b) Determine if the current segment is in deep water (at least N sectors project over water that is at least "depth threshold" meters)

    (c) Determine if the current segment is enclosed by land (at least N fetch rays have to be blocked by land, i.e. fetch distance is less than "ocean effect cutoff" meters).

    (d) Determine segment exposure: a shore segment is exposed if it is both in deep waters, and not enclosed by land (facing open water), otherwise, it is sheltered.

In other words, if the fetch threshold is 12 km and the depth threshold is 5 m, and the exposure proportion is 0.8, the model will classify a segment as sheltered if less than 80% of the segments have a fetch distance lower than 12 km *or* the average depth along each fetch segment is less than 5 m.

Name: A numeric text string (number between 0 and 1) File type: text string (direct input to the interface) Sample (default): 0.8

1. **Oceanic effect cutoff (meters, required).** Used as a threshold to determine if a shore segment is enclosed by land. See Exposure proportion, step 3.

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the interface)
Sample (default): 60000
```

2. **Geomorphology: Shoreline Type (optional).** This input is used to compute the Geomorphology ranking of each shoreline segment (*Table 4.1*). It does not have to match the land polyline input, but must resemble it as closely as possible. If it doesn't, the model will try to match the coastlines using the coastal overlap parameter. Additionally, the polyline shapefile must have a field called "RANK" that identifies the various shoreline type ranks with a number from 1-5. More information on how to fill in this table is provided in *Appendix B*.

```
Names: File can be named anything, but no spaces in the name
File type: polyline shapefile (.shp)
Sample path: \InVEST\CoastalProtection\Input\Geomorphology_BarkClay.shp
```

3. **Layer value if path omitted (optional).** Integer value between 1 and 5. If geomorphology is omitted, replace all shore points for this layer with a constant rank value in the computation of the coastal vulnerability index. If both the file and value for the layer are omitted, the layer is skipped altogether.

```
Name: A positive integer between 1 and 5.
File type: text string (direct input to the interface)
Sample (default): empty
```

4. **Coastal overlap (meters, required).** Tolerance threshold in meters (that should be a multiple of cell size), to make 2 non-overlapping shorelines match. If the tolerance is twice the cell size, the model will be able to match shorelines that are 2 pixels off. If it is 4 times the cell size, the model will be able to match shorelines that are 4 pixels off, and so on. It's used when the shoreline from geomorphology doesn't overlap exactly with the shoreline from the land polygon shapefile.

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the interface)
Sample (default): 250
```

5. **Natural Habitat (optional).** Directory that contains habitat layers. Users must store all Natural Habitats input layers that they have in a unique directory. The model uses these input layers to compute a Natural Habitat ranking for each shoreline segment. All data in this directory must be polyline or polygon shapefiles that depict the location of the habitats, and must be projected in meters. Additionally, each layer must end with an underscore followed by a unique alpha-numeric number. The model uses that number to match the habitat layer to the information that users provide in the CSV table (see next input). The model allows for a maximum of eight layers in this directory. Do not store any additional files that are not part of the analysis in this directory. The distance at which this layer will have a protective influence on coastline can be modified in the natural habitat CSV table (input 8).

```
Name: Folder can be named anything, but no spaces in the name.  Habitat layers in the folder mus
File type:a polygon shapefile (.shp)
Sample path: \InVEST\CoastalProtection\Input\NaturalHabitat
```

6. **Natural Habitat Layers CSV (Table optional).**. Users must provide a summary table to instruct the model on the protective influence (rank) and distance of natural habitat. Use the sample table provided as a template since the model expects values to be in these specific cells. More information on how to fill this table is provided in *Appendix B*.

```
Table Names: File can be named anything, but no spaces in the name
File type: *.csv
Sample: InVEST\CoastalProtection\Input\NaturalHabitat_WCVI.csv
```

	A	B	C	D
1	HABITAT	ID	RANK	PROTECTION DISTANCE (m)
2	kelp	1	4	1500
3	eelgrass	2	4	500
4	high dune	3	2	300
5	low dune	4	3	300
6				

22. **Layer value if path omitted (optional).** Integer value between 1 and 5. If natural habitats is omitted, replace all shore points for this layer with a constant rank value in the computation of the coastal vulnerability index. If both the file and value for the layer are omitted, the layer is skipped altogether.

```
Name: A positive integer between 1 and 5.
File type: text string (direct input to the interface)
Sample (default): empty
```

23. **Climatic forcing grid (optional).** This input is used to compute the Wind and Wave Exposure ranking of each shoreline segment (*Table 4.1*). It consists of a point shapefile that contains the location of the grid points as well as wave and wind values that represent storm conditions at that location. If users would like to create such a file from their own data, instructions are provided in *Appendix B*.

```
Name: File can be named anything
Format: point shapefile where each point has information about wind and wave measurements.
Sample data set (default): \InVEST\CoastalProtection\Input\WaveWatchIII.shp
```

24. **Layer value if path omitted (optional).** Integer value between 1 and 5. If climatic forcing grid is omitted, replace all shore points for this layer with a constant rank value in the computation of the coastal vulnerability index. If both the file and value for the layer are omitted, the layer is skipped altogether.

```
Name: A positive integer between 1 and 5.
File type: text string (direct input to the interface)
Sample (default): empty

Sample (default): 500
```

25. **Continental Shelf (optional).** This input is a global polygon dataset that depicts the location of the continental margin. It must intersect with the AOI polygon (input #2).

```
Names: File can be named anything, but no spaces in the name
File type: polygon shapefile (.shp)
Sample path:  \InVEST\CoastalProtection\Input\continentalShelf.shp
```

26. **Depth contour level (meters, optional).** If no continental shelf is specified, the model will use the bathymetry data to trace a user-defined depth contour level ans use it instead of the edge of the continental shelf.

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the interface)
Sample (default): 150
```

27. **Sea Level Rise (optional).** Polygon Indicating Net Rise or Decrease. This input must be a polygon delineating regions within the AOI that experience various levels of net sea level change. It must have a field called "Trend", which represents the rate of increase (mm/yr) of the sea level in a particular region according to *Table 4.1*. More information on how to create this polygon is provided in *Appendix B*. For general help with creating and editing shapefiles, try documentation provided by ArcGIS or QGIS.:

```
Name: File can be named anything, but no spaces in the name
File type: polygon shapefile (.shp) or point shapefile (.shp)
Sample path: \InVEST\CoastalProtection\Input\SeaLevRise_WCVI.shp
```

28. **Layer value if path omitted (optional).** Integer value between 1 and 5. If sea level rise is omitted, replace all shore points for this layer with a constant rank value in the computation of the coastal vulnerability index. If both the file and value for the layer are omitted, the layer is skipped altogether.

```
Name: A positive integer between 1 and 5.
File type: text string (direct input to the interface)
Sample (default): empty
```

29. **Structures (optional).** Polygon shapefile that contains the location of rigid structures along the coast.

```
Name: File can be named anything, but no spaces in the name
File type: polygon shapefile (.shp)
Sample path: \InVEST\CoastalProtection\Input\Structures_BarkClay.shp
```

30. **Layer value if path omitted (optional).** Integer value between 1 and 5. If structures layer is omitted, replace all shore points for this layer with a constant rank value in the computation of the coastal vulnerability index. If both the file and value for the layer are omitted, the layer is skipped altogether.

```
Name: A positive integer between 1 and 5.
File type: text string (direct input to the interface)
Sample (default): empty
```

31. **Population Raster (optional).** If provided, a raster grid of population is used to map the population size along the coastline of the AOI specified (input #4). A global population raster file is provided as default, but other population raster layers can be substituted.

```
Name: File can be named anything, but no spaces in the name and less than 13 characters
Format: standard GIS raster file (ESRI GRID), with population values
Sample data set (default): \InVEST\Base_Data\Marine\Population\global_pop\hdr.adf
```

32. **Min. population in urban centers (required).** Minimum population that has to live in the vincinity of a shore segment to be considered a urban center. The vincinity is defined in the next input, "coastal neighborhood".

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the interface)
Sample (default): 5000
```

33. **Coastal neighborhood (radius in m, required).** Radius defining the vincinity of a shore segment that is used to count the population living on or near the coast.

```
Name: A numeric text string (positive integer)
File type: text string (direct input to the interface)
Sample (default): 150
```

34. **Additional layer (optional).** This additional layer can be any additional variable that users desire to add to the exposure index. It can be values of long-term shoreline change, for example. This layer must be a polygon shapefile with a trend column that separates different shoreline regions characteristics in its attribute table. Once all segments have a value, we rank them according to quartile distribution.

```
Name: File can be named anything, but no spaces in the name and less than 13 characters
Format: standard GIS raster file (ESRI GRID), with population values
Sample data set (default): \InVEST\Base_Data\Marine\Population\global_pop
```

35. **Layer value if path omitted (optional).** Integer value between 1 and 5. If additional layer is omitted, replace all shore points for this layer with a constant rank value in the computation of the coastal vulnerability index. If both the file and value for the layer are omitted, the layer is skipped altogether.

```
Name: A positive integer between 1 and 5.
File type: text string (direct input to the interface)
Sample (default): empty
```

6.2.6 Running the Model

Setting up Workspace and Input Directories

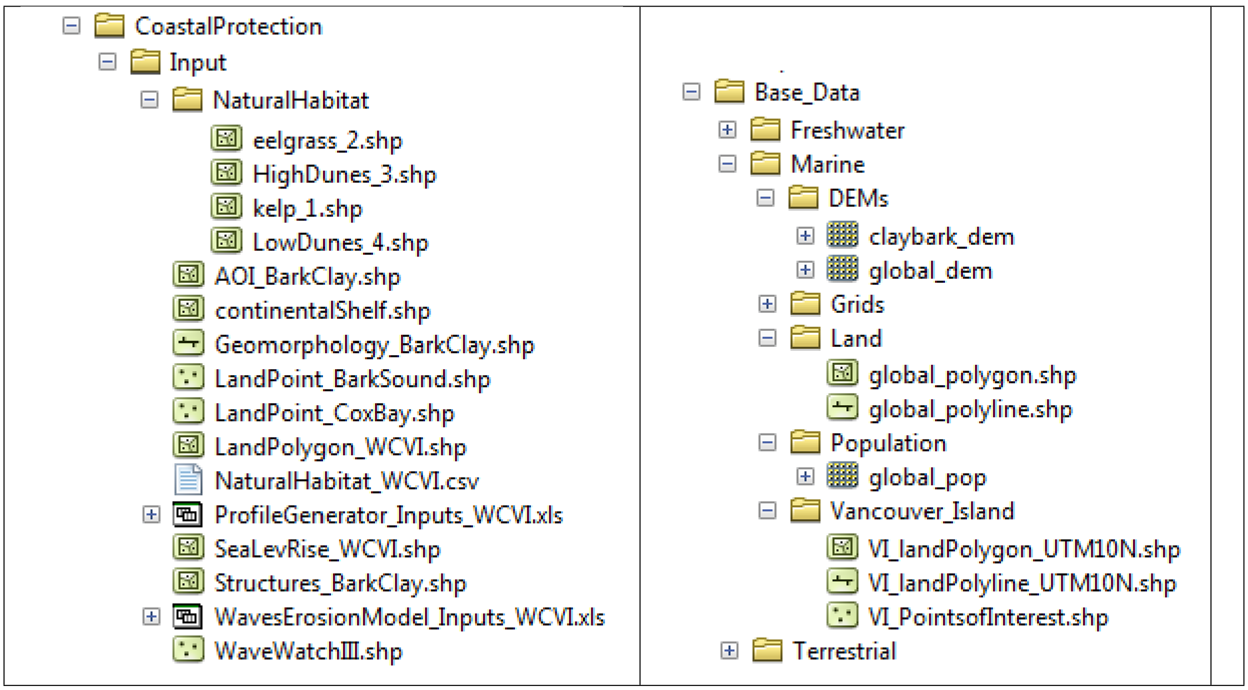
These directories will hold all input, intermediate and output data for the model.

Note: The word *'path'* means to navigate or drill down into a directory structure using the Open Folder dialog window that is used to select GIS layers or Excel worksheets for model input data or parameters.

Exploring a project workspace and input data directory

The */InVEST/CoastalProtection* directory holds the main working directory for the model and all other associated directories. Within the *CoastalProtection* directory there will be a sub-directory named *'Input'*. This directory holds most of the GIS and tabular data needed to setup and run the model.

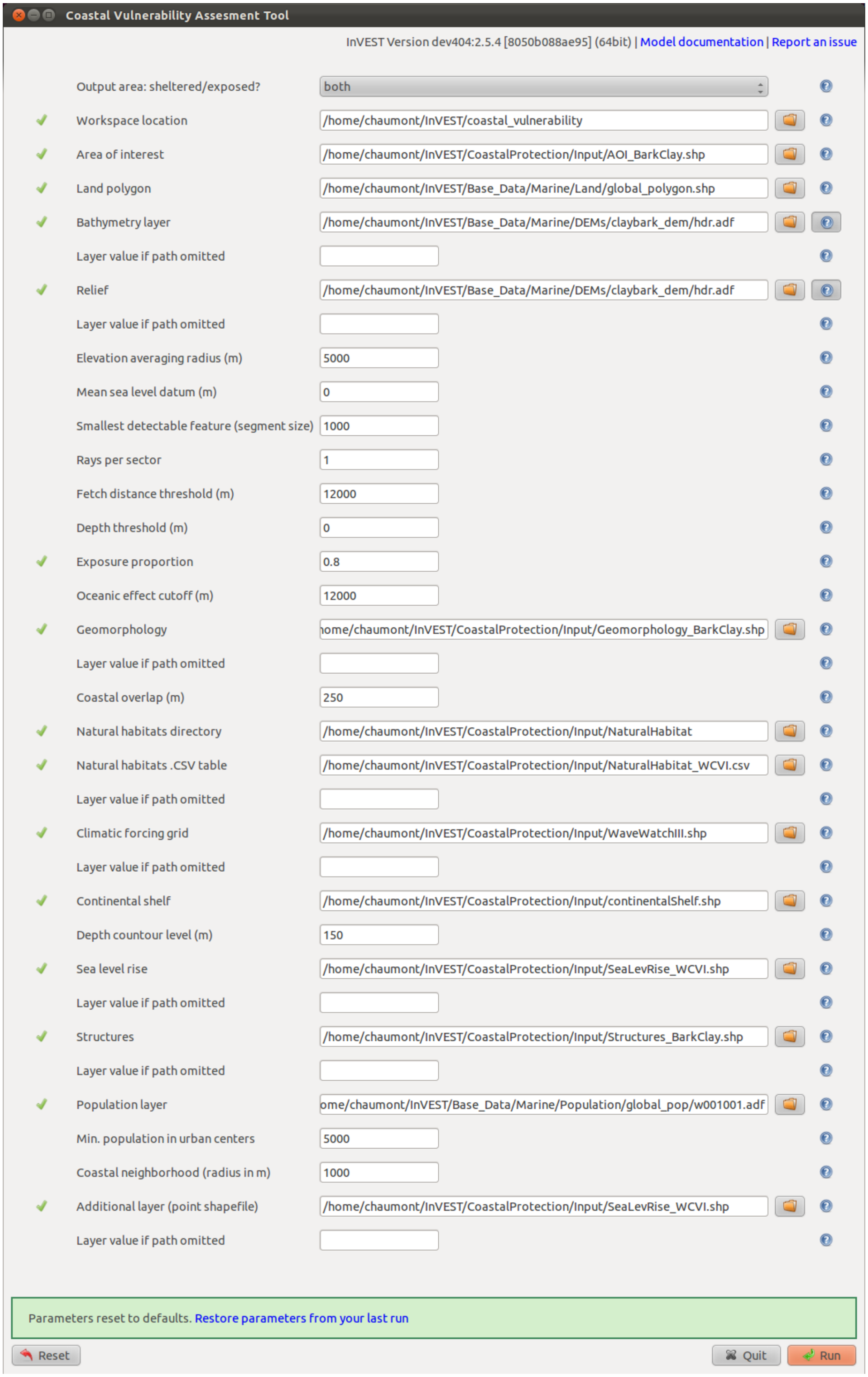
The following image shows the sample input (on the left) and base data (on the right) directory structures and accompanying GIS data. It is recommended that this directory structure be used as a guide to organize workspaces and data. Refer to the screen-shots below for examples of directory structure and data organization.



Creating a run of the model

The following example of setting up the Coastal Vulnerability model uses the sample data provided with the InVEST download. The instructions and screen-shots refer to the sample data and directory structure supplied within the InVEST installation package. We expect that users will have location-specific data to use in place of the sample data. These instructions provide only a guideline on how to specify to the standalone 3.0 version of the model the various types of data needed and does not represent any site-specific model parameters. See the *Data Needs* section for a more complete description of the data specified below.

1. On Windows 7, click on the "start" button to expand the start menu, then click on "All Programs" at the bottom.
2. Expand the folder which name starts with "InVEST", and launch the model by clicking on "Coastal Vulnerability". The model will show a user interface as shown in the next page.
3. Specify the area to appear in the output: sheltered shoreline segments, or both sheltered and exposed.
4. Specify the Workspace. Either enter the path to the workspace manually (the model will create it if it doesn't already exist), or click on it from the navigation window (click on the Open Folder button on the right, default


Coastal Vulnerability Assesment Tool
InVEST Version dev404:2.5.4 [8050b088ae95] (64bit) | Model documentation | Report an issue
Output area: sheltered/exposed?
both
Workspace location
/home/chaumont/InVEST/coastal_vulnerability
Area of interest
/home/chaumont/InVEST/CoastalProtection/Input/AOI_BarkClay.shp
Land polygon
/home/chaumont/InVEST/Base_Data/Marine/Land/global_polygon.shp
Bathymetry layer
/home/chaumont/InVEST/Base_Data/Marine/DEMs/claybark_dem/hdr.adf
Layer value if path omitted
Relief
/home/chaumont/InVEST/Base_Data/Marine/DEMs/claybark_dem/hdr.adf
Layer value if path omitted
Elevation averaging radius (m)
5000
Mean sea level datum (m)
0
Smallest detectable feature (segment size)
1000
Rays per sector
1
Fetch distance threshold (m)
12000
Depth threshold (m)
0
Exposure proportion
0.8
Oceanic effect cutoff (m)
12000
Geomorphology
ome/chaumont/InVEST/CoastalProtection/Input/Geomorphology_BarkClay.shp
Layer value if path omitted
Coastal overlap (m)
250
Natural habitats directory
/home/chaumont/InVEST/CoastalProtection/Input/NaturalHabitat
Natural habitats .CSV table
/home/chaumont/InVEST/CoastalProtection/Input/NaturalHabitat_WCVI.csv
Layer value if path omitted
Climatic forcing grid
/home/chaumont/InVEST/CoastalProtection/Input/WaveWatchIII.shp
Layer value if path omitted
Continental shelf
/home/chaumont/InVEST/CoastalProtection/Input/continentalShelf.shp
Depth countour level (m)
150
Sea level rise
/home/chaumont/InVEST/CoastalProtection/Input/SeaLevRise_WCVI.shp
Layer value if path omitted
Structures
/home/chaumont/InVEST/CoastalProtection/Input/Structures_BarkClay.shp
Layer value if path omitted
Population layer
ome/chaumont/InVEST/Base_Data/Marine/Population/global_pop/w001001.adf
Min. population in urban centers
5000
Coastal neighborhood (radius in m)
1000
Additional layer (point shapefile)
/home/chaumont/InVEST/CoastalProtection/Input/SeaLevRise_WCVI.shp
Layer value if path omitted
Parameters reset to defaults. Restore parameters from your last run
Reset
Quit
Run

is *InVEST/CoastalProtection*). This is the directory in which the intermediate and final outputs will be stored when the model is run.

5. Specify the Area of Interest (AOI). The model requires an AOI, which is the geographic area over which the model will be run. This example refers to the sample shapefile *AOI_BarkClay.shp* supplied in *InVEST/CoastalProtection/Input*. For general help with creating and editing shapefiles, try documentation provided by ArcGIS or QGIS. For more specific InVEST-related GIS video tutorials, consider enrolling in the online course, Introduction to the Natural Capital Project Approach.
6. Specify the Land Polygon. The model requires a land polygon shapefile to define the shoreline for the analysis. A default path to the global sample data is supplied in the model window for users.
7. Specify the bathymetry (DEM raster) of the water in the AOI to be incorporated into Wave Exposure calculations. It will be used to estimate wave height and associated period, for each of the 16 fetch angular sectors.
8. Specify the Relief Digital Elevation Model (DEM) raster. The model requires a DEM raster file to estimate average elevation landward of the coastal segment. The path of the default DEM file for the west coast of Vancouver Island is in *InVEST/Base_Data/Marine/DEMs/claybark_dem*.
8. Specify the elevation averaging radius (default is 5000, i.e. 5km). The model will average all the land elevations within this radius to compute relief.
9. Specify the mean sea level datum. The model can adjust the mean sea level relative to the datum of the bathymetry. Default is zero, and positive values indicate that the mean sea level datum is above the bathymetry's datum.
10. Specify the size of the smallest detectable feature (Cell Size). The model requires a cell size for the raster analysis (default is 250 m). The model will not be able to distinguish details on rasters or shapefiles that are smaller than the size specified.
11. Specify the number of rays per sector: This is the number of rays the model uses to sample water depth and the presence of landmasses within each sector. Users may change this value by entering a new value directly into the text box.
12. Specify the maximum fetch distance. The model computes the fetch over a maximum distance to separate sheltered and exposed areas. The default value is 12,000 meters. The longer the distance, the longer the rays, and the slower the computation.
13. Specify a depth threshold (positive integer, or 0). The model uses the depth threshold to determine areas of shallow water. It is used to segregate exposed from sheltered shore segments. A value of zero (0) cancels the effect of this parameter.
14. Specify exposure proportion (real number between 0 and 1). The model requires a percentage of sectors to span either shallow or enclosed water bodies to classify the shore as sheltered. uses the depth threshold to determine areas of shallow water. It is used to segregate exposed from sheltered shore segments. A value of zero (0) cancels the effect of this parameter, as all water bodies will be considered deep.
15. The Oceanic effect cutoff is the maximum distance allowed to consider coastal segments enclosed. Set this value so that it is at least as long as the distance across areas that should be enclosed.
16. Specify the Geomorphology layer (optional). The model can use an optional polygon shapefile that represents shoreline geomorphology.
17. Coastal overlap should be a multiple of cell size. This multiple is the maximum number of pixels between two non-overlapping shorelines the model can cope with when processing geomorphology. If the geomorphology comes from local data and the shoreline is from a global dataset, the coastlines might not overlap completely. To resolve the discrepancy, increase coastal overlap to several multiples of cell size, otherwise, leave it at zero.
18. Specify the Natural Habitat directory (optional). The model can use optional polygon shapefiles that represent the location of various habitats.

19. Specify the Natural Habitat CSV table (optional). If the above input for natural habitat directory is specified, the model requires this table of habitat ranks and protective distance stored in a CSV. See the *Data Needs* section for more information on creating and formatting this table. A sample CSV will is supplied.
20. Specify the climatic forcing grid (Wind-Wave data point shapefile). The model requires wind and wave statistics to create the wind and wave exposure variables. See the *Data Needs* section for details on preparing a shapefile from another data source.
21. Specify the continental shelf layer (optional). To represent surge potential, the model uses a continental shelf polygon shapefile.
22. Specify the Sea Level Rise layer (optional). The model can use an optional polygon shapefile that represents sea level rise potential.
23. Specify the structures layer. Polygon shapefile that contains the location of rigid structures along or near the coast.
24. Specify the population layer. This file should be a raster population assigned to each cell value. The default data for this layer is a global raster located in *InVEST/Base_Data/Marine/Population/global_pop*. If users have a superior raster, they are instructed to select the location of this data on their local computer.
25. Set the minimum population threshold in urban centers (default is 5000 people). The model will sum the population within a user-defined radius (see coastal neighborhood below) and will report segments that exceed the threshold.
26. Set the coastal neighborhood (default is 2000 meters). The model will sum the population within the specified radius. It is important to keep in mind that the surface over which the population is aggregated increases as the square of the radius.
27. At this point the model dialog box is completed for a complete run (with all optional data for full exposure analysis) of the Coastal Vulnerability model.

    Click the "run" button at the lower right corner of the window to run the model. A new window will appear and show the progress about each step in the analysis. It will also show the most salient warnings when preprocessing the input, as well as warnings during the computation of the various indices. Once the model finishes, the progress window will show all the completed steps and the amount of time that has elapsed during the model run.

Upon successful completion of the model, two new directories called "intermediate" and "outputs" will be created in the workspace. The main outputs that are useful for further analysis are in the "coastal_vulnerability" and "population" sub-directories in "outputs". The remainder of this guide will concentrate on these outputs. The types of spatial data that is generated are described in the cv-interpreting-results section.

6.2.7 Interpreting results

Model outputs

The following is a short description of each of the outputs from the Coastal Vulnerability model. Files are grouped in sub-directories within the "intermediate/" and "outputs/" directories, except for "vulpes.sefs.uw.edu-ttapp-cv-dash.php" that is directly located in "outputs/". The "outputs/" and "intermediate/" directories are saved in the workspace directory that was specified by the user. Every sub-directory has a comma separated file (CSV) that is a text version of the compiled data of all raster files present in the subdirectory for each shoreline segment. Depending on the first option in the model's user interface, these shore segments either cover the sheltered areas only, or both the sheltered and exposed ones.

Output directory

- outputs\vulpes.sefs.uw.edu-ttapp-cv-dash.php
  - This file is a shortcut to an html page for a coastal vulnerabilty dashboard that allows users to upload their 'coastal_exposure.csv'. Then through an intractive map, different layer results can be viewed and interpreted.
- outputs\coastal_exposure: contains all the layers used to compute the coastal vulnerability index.
  - 1_a_shore_exposure.tif - a raster where the cells corresponding to the shoreline segments are either 0 if sheltered or 1 if exposed.
  - 1_b_geomorphology.tif - a raster where shore segments are valued from 1 to 5 depending on the geomorphology in the geomorphology layer. Lower coastal values indicate geomorphologic types are less susceptible to erosion, and vice-versa.
  - 1_c_relief.tif - a raster where shore segments are valued from 1 to 5 depending on the average elevation around that cell. Lower values indicate lower elevations.
  - 1_d_natural_habitats.tif - a raster where shore segments are valued according to the natural habitats that are present there. The model uses equation (6.6) that uses natural habitat ranks specified in *Table 4.1*. according to their exposure to winds.
  - 1_f_wave_exposure.tif - a raster where shore segments are ranked in a similar way to wind exposure, but according to their exposure to wave.
  - 1_g_surge_potential.tif - a raster where segments are ranked according to their exposure to potential surge. First, the exposed segments are assigned a rank in equal proportion between 1 and 5, depending on their distance to the edge of the continental shelf. Then, these values are propagated along the sheltered coast. Isolated coastline segments (such as islands) are assigned the rank of the closest (already ranked) segment.
  - 1_h_sea_level_rise.tif - a raster with segments ranked in equal proportion between 1 and 5 based on the sea level rise value from the input shapefile.
  - 1_i_coastal_exposure.tif - a raster with the coastal exposure index computed as in (6.3).
  - 1_j_coastal_exposure_no_habitats.tif - raster containing values computed from the same equation as the coastal exposure raster except the natural habitats layer has been replaced by the constant 5.
  - 1_k_habitat_role.tif - raster difference between coastal_exposure_no_habitats and coastal_exposure.
  - 3_2_erodible_shoreline.tif - raster where the shoreline segment values are computed with equation (6.5).
  - coastal_exposure.csv - comma-separated file that aggregates the data in each file in the directory for each coastal segment
  - coastal_exposure.shp - a Point Shapefile that plots the data from 'coastal_exposure.csv'.
- outputs\population: contains all the layers used to compute the coastal vulnerability index.
  - 0_structures_edges.tif - a raster with only the shore segments that border the coastal structures.
  - 1_a_shore_exposure.tif - same as in the "coastal_vulnerability/" sub-directory.
  - 1_i_coastal_vulnerability.tif - same as in the "coastal_vulnerability/" sub-directory.
  - 1_j_coastal_vulnerability_no_habitats.tif - same as in the "coastal_vulnerability/" subdirectory.
  - 1_m_coastal_population.tif - raster where every coastal segment having the population living on the coast.

Intermediate directory

The model currently generates several hundreds of files classified in sub-directories in the intermediate directory. There is one intermediate subdirectory per computational step required to produce a file in the "outpus/" directory. Each of these sub-directory is prefixed so that the alphabetical order reflects the order of the model's computational steps. Within a sub-directory, each file is the result of a computational step and is usually numbered so that it is possible to follow the order the computation that is carried out. If there is a problem with an output file, the user can go back to the corresponding sub-directory and look at the intermediate files individually to infer what happened during the computation.

Intermediate folders also contain raw outputs for each of the variable in separate .csv files. Users can also use the .csv files in each of the intermediate folders to post-process data and generate new outputs, using, e.g., a different formulation for the exposure index or a different ranking system for each variable.

Parameter log

Each time the module is run a text file will appear in the workspace directory. The file will list the parameter values for that run and be named according to the service and the date and time.

6.2.8 Appendix A

In this appendix, definitions for the terms presented in the geomorphic classification in *Table 4.1* are presented. Some of these are from Gornitz et al. (1997) and USACE (2002). Photos of some of the geomorphic classes that are presented can be found at the National Oceanic and Atmospheric Administration's Ocean Service Office of Response and Restoration website.

Alluvial Plain A plain bordering a river, formed by the deposition of material eroded from areas of higher elevation.

Barrier Beach Narrow strip of beach with a single ridge and often foredunes. In its most general sense, a barrier refers to accumulations of sand or gravel lying above high tide along a coast. It may be partially or fully detached from the mainland.

Beach A beach is generally made up of sand, cobbles, or boulders and is defined as the portion of the coastal area that is directly affected by wave action and that is terminated inland by a sea cliff, a dune field, or the presence of permanent vegetation.

Bluff A high, steep backshore or cliff

Cliffed Coasts Coasts with cliffs and other abrupt changes in slope at the ocean-land interface. Cliffs indicate marine erosion and imply that the sediment supply of the given coastal segment is low. The cliff's height depends upon the topography of the hinterland, lithology of the area, and climate.

Delta Accumulations of fine-grained sedimentary deposits at the mouth of a river. The sediment is accumulating faster than wave erosion and subsidence can remove it. These are associated with mud flats and salt marshes.

Estuary Coast The tidal mouth of a river or submerged river valley. Often defined to include any semi-enclosed coastal body of water diluted by freshwater, thus includes most bays. The estuaries are subjected to tidal influences with sedimentation rates and tidal ranges such that deltaic accumulations are absent. Also, estuaries are associated with relatively low-lying hinterlands, mud flats, and salt marshes.

Fiard Glacially eroded inlet located on low-lying rocky coasts (other terms used include sea inlets, fjard, and firth).

Fjord A narrow, deep, steep-walled inlet of the sea, usually formed by the entrance of the sea into a deep glacial trough.

Glacial Drift A collective term which includes a wide range of sediments deposited during the ice age by glaciers, melt-water streams and wind action.

Indented Coast Rocky coast with headland and bays that is the result of differential erosion of rocks of different erodibility.

Lagoon A shallow water body separated from the open sea by sand islands (e.g., barrier islands) or coral reefs.

Mud Flat A level area of fine silt and clay along a shore alternately covered or uncovered by the tide or covered by shallow water.

6.2.9 Appendix B

The model requires large-scale geophysical, biological, atmospheric, and population data. Most of this information can be gathered from past surveys, meteorological and oceanographic devices, and default databases provided with the model. In this section, various sources for the different data layers that are required by the model are proposed, and methods to fill out the input interface discussed in the *Data Needs* section are described. It is recommend that users import all the required and optional data layers before attempting to run the model. Familiarity with data layers will facilitate the preparation of data inputs.

Population data

To assess the population residing near any segment of coastline, population data from the Global Rural-Urban Mapping Project (GRUMP) is used. This dataset contains global estimates of human populations in the year 2000 in 30 arc-second (1km) grid cells. User are encouraged to use their own, more detailed and/or recent census data, and it is encouraged that recent fine-scale population maps are used, even in paper form, to aid in the interpretation of the Exposure Index map.

Geo-physical data layer

To estimate the Exposure Index of the AOI, the model requires an outline of the coastal region. As mentioned in the *Data Needs* Section, we provide a default global land mass polygon file. This default dataset, provided by the U.S. National Oceanic and Atmospheric Administration (NOAA) is named GSHHS, or a Global Self-consistent, Hierarchical, High-resolution Shoreline (for more information, visit http://www.ngdc.noaa.gov/mgg/shorelines/gshhs.html). It should be sufficient to represent the outline of most coastal regions of the world. However, if this outline is not sufficient, we encourage that users substitute it with another layer.

To compute the Geomorphology ranking, users must provide a geomorphology layer (*Data Needs* Section, input 15) and an associated geomorphic classification map. This map should provide the location and type of geomorphic features that are located in the coastal area of interest. In some parts of the west-coast of the United States and Canada, such a map can be built from a database called Shorezone. For other parts of the United States, users can consult the Environmental Sensitivity Index website. If such a database is not available, it is recommend that a database from site surveys information, aerial photos, geologic maps, or satellites images (using Google or Bing Maps, for example) is built. State, county, or other local GIS departments may have these data, freely available, as well.

In addition to the geomorphology layer, users must have a field in its attribute table called "RANK". This is used by the model to assign a geomorphology exposure ranking based on the different geomorphic classes identified. Assign the exposure ranks based on the classification presented in *Table 4.1*. All ranks should be numeric from 1 to 5.

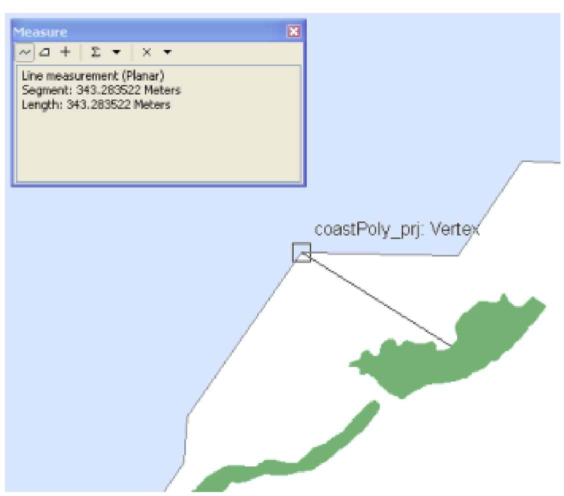
Habitat data layer

The natural habitat maps (inputs 7 and 8 in the *Data Needs* Section) should provide information about the location and types of coastal habitats described in *Table 4.1*. The subtidal layers in that directory have been built from a database called Shorezone. Dune data from unpublished an dataset provided by Raincoast Applied Ecology was obtained. If such a database is not available, it is recommend building it from site surveys information, aerial photos, or even satellites images (using Google or Bing Maps, for example).

The Natural Habitat CSV table input asks users to provide information about the type of habitats layers that users have in the "NaturalHabitat" directory. The different columns in that table are:

1. HABITAT: The name of the natural habitat for which users have a layer (e.g., kelp or eelgrass)
2. ID: The ID number associated with the name of these habitats: the underscored integer number X listed at the end of the name of the different layers that have been created, as in "eelgrass_2". Note that this ID number is what the model uses to associate a rank and protection distance to the name users input in the first column. In other words, the name in column 1 can be different from the name of your file, but the ID number should match. For example, in the default natural habitat layers directory that has been provided, the eelgrass layer has the ID = 2 (e.g. eelgrass_2). Since the ID in the second column is 2, then the model recognizes that the rank and protection distance values that are defined for "eelgrass" apply to the eelgrass_2.shp layer.
3. RANK: The vulnerability rank associated with the natural habitat that is listed in column 1. It is recommend that the ranking system provided in *Table 4.1* is used. However, if users would like to evaluate how the vulnerability index values changes in the absence of the habitats listed in the table, users should change the RANK to a 5. For example, to evaluate how the vulnerability of an area changes if high sand dunes are removed, users can change the RANK value for high sand dunes from a 2 to a 5.
4. PROTECTIVE DISTANCE (m): The model determines the presence or absence of various natural habitats that users specified in the AOI by estimating the fetch distance over the 16 equiangular segments between the location of the natural habitats and the shoreline. If there is a non-zero fetch distance between a patch of natural habitat and a shoreline segment, the model recognizes that the patch fronts that segment. To assign a natural habitat ranking to that segment which takes into account the beneficial effect of the presence of this habitat, it is asked that users input a maximum distance of influence into the Natural Habitat CSV table (input 8). It is assumed that natural habitats that are fronting a segment but are further away from the segment than the distance that is defined by the user will not have a beneficial effect on the stability of that segment, and will not be counted in the natural habitat ranking for that segment.

To estimate this distance, it is recommend that users load the various habitat layers located in their "Natural Habitats" directory as well as the polygon layer representing the area of interest. Then, using the ArcGIS "distance" tool, measure the distance between the shoreline and natural habitats that you judge to be close enough to have an effect on nearshore coastal processes. It is best to take multiple measurements and develop a sense of an average acceptable distance that can serve as input. Please keep in mind that this distance is reflective of the local bathymetry conditions (a seagrass bed can extend for kilometers seaward in shallow nearshore regions), but also of the quality of the spatial referencing of the input layer. The example below gives an example of such measurement when seagrass beds are considered (green patches).



As mentioned in *Natural Habitats*, the model computes the natural habitat exposure ranking for a shoreline segment using the following equation:

$$R_{Hab} = 4.8 - 0.5\sqrt{(1.5\max_{k=1}^{N}(5-R_k))^2 + (\sum_{k=1}^{N}(5-R_k)^2 - \max_{k=1}^{N}(5-R_k))^2)}$$

This equation is applied to various possible combinations of natural habitats, and the results of this exercise are presented in the table and figure below:

Wind data

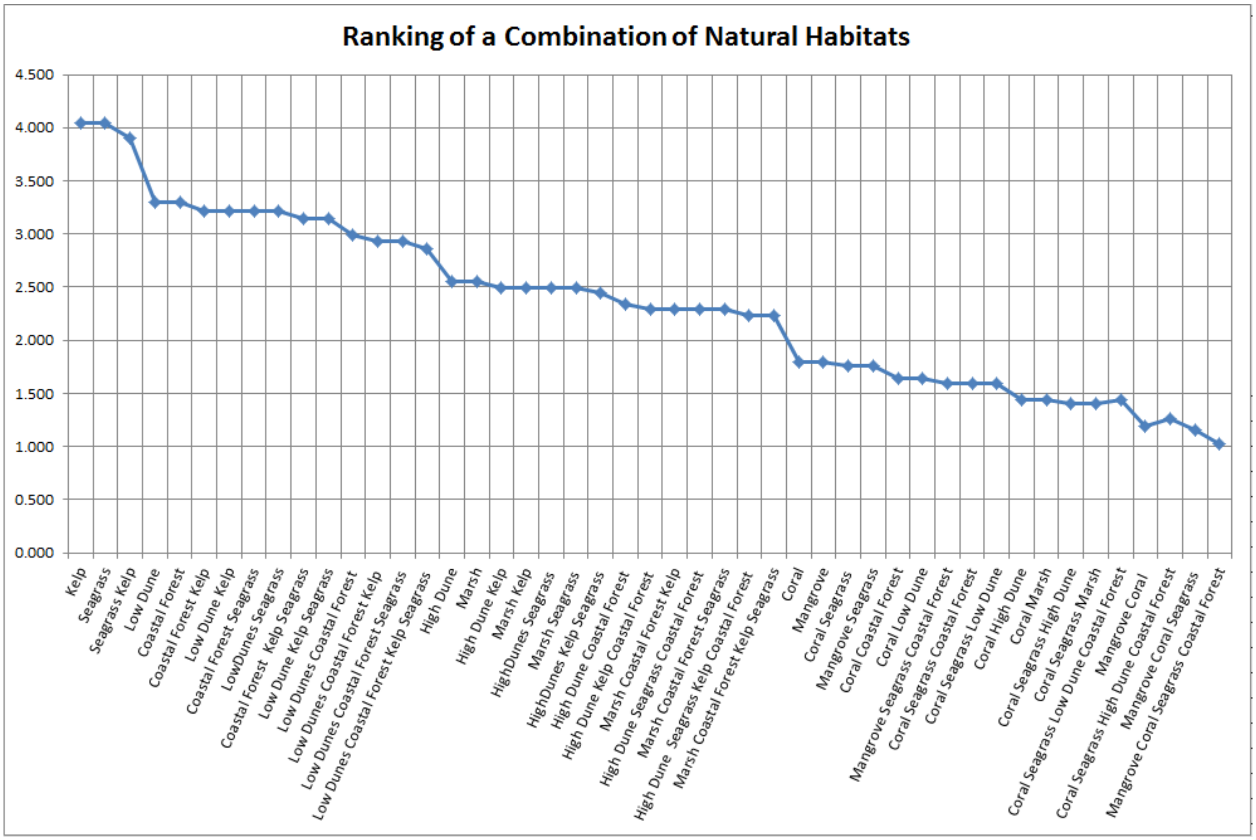
To estimate the importance of wind exposure and wind-generated waves, wind statistics measured in the vicinity of the AOI are required. From at least 5 years of data, the model requires the average in each of the 16 equiangular sectors (0deg, 22.5deg, etc.) of the wind speeds in the 90th percentile or greater observed near the segment of interest to compute the REI. In other words, for computation of the REI, sort wind speed time series in descending order, and take the highest 10% values, and associated direction. Sort this sub-series by direction: all wind speeds that have a direction centered around each of the 16 equiangular sectors are assigned to that sector. Then take the average of the wind speeds in each sector. If there is no record of time series in a particular sector because only weak winds blow from that direction, then average wind speed in that sector is assigned a value of zero (0). Please note that, in the model, wind direction is the direction winds are blowing FROM, and not TOWARDS.

For the computation of wave power from wind and fetch characteristics, the model requires the average of the wind speeds greater than or equal to the 90th percentile observed in each of the 16 equiangular sectors (0deg, 22.5deg, etc.). In other words, for computation of wave power from fetch and wind, sort the time series of observed wind speed by direction: all wind speeds that have a direction centered on each of the 16 equiangular sectors are assigned to that sector. Then, for each sector, take the average of the highest 10% observed values. Again, please note that, in our model, wind direction is the direction winds are blowing FROM, and not TOWARDS.

If users would like to provide their own wind and wave statistics, instead of relying on WW3 data, the must enter the data in the following order:

1. Column 1-2: Placeholder. No information required.
2. Columns 3-4: LAT, LONG values. These values indicate the latitude and longitude of the grid points that will be used to assign wind and wave information to the different shoreline segments.
3. Columns 5-20: REI_VX, where X=[0,22,45,67,90,112,135,157,180,202,225,247,270,292,315,337] (e.g., REI_V0). These wind speed values are computed to estimate the REI of each shoreline segment. These values are the average of the highest 10% wind speeds that were allocated to the 16 equiangular sectors centered on the angles listed above.
4. Columns 21 to 36: REI_PCTX, where X has the same values as listed above. These 16 percent values (which sum to 1 when added together) correspond to the proportion of the highest 10% wind speeds which are centered on the main sector direction X listed above.
5. Column 37 to 52: WavP_X, where X has the same values as listed above. These variables are used to estimate wave exposure for sites that are directly exposed to the open ocean. They were computed from WW3 data by first estimating the wave power for all waves in the record, then splitting these wave power values into the 16 fetch sectors defined earlier. For each sector, we then computed WavP by taking the average of the top 10% values (see Section *The Model*).
6. Column 53 to 68: WavPPCTX, where X has the same values as listed above. These variables are used in combination with WavP_X to estimate wave exposure for sites that are directly exposed to the open ocean. They correspond to the proportion of the highest 10% wave power values which are centered on the main sector direction X (see Section *The Model*).
7. Columns 69 to 84: V10PCT_X, where X has the same values as listed above. These variables are used to estimate wave power from fetch. They correspond to the average of the highest 10% wind speeds that are centered on the main sector direction X.

Combination of Natural Habitats	Final Ranking
Kelp	4.050
Seagrass	4.050
Seagrass Kelp	3.899
Low Dune	3.300
Coastal Forest	3.300
Coastal Forest Kelp	3.219
Low Dune Kelp	3.219
Coastal Forest Seagrass	3.219
LowDunes Seagrass	3.219
Coastal Forest Kelp Seagrass	3.142
Low Dune Kelp Seagrass	3.142
Low Dunes Coastal Forest	2.997
Low Dunes Coastal Forest Kelp	2.929
Low Dunes Coastal Forest Seagrass	2.929
Low Dunes Coastal Forest Kelp Seagrass	2.864
High Dune	2.550
Marsh	2.550
High Dune Kelp	2.495
Marsh Kelp	2.495
HighDunes Seagrass	2.495
Marsh Seagrass	2.495
HighDunes Kelp Seagrass	2.442
High Dune Coastal Forest	2.338
High Dune Kelp Coastal Forest	2.288
Marsh Coastal Forest Kelp	2.288
High Dune Seagrass Coastal Forest	2.288
Marsh Coastal Forest Seagrass	2.288
High Dune Seagrass Kelp Coastal Forest	2.238
Marsh Coastal Forest Kelp Seagrass	2.238
Coral	1.800
Mangrove	1.800
Coral Seagrass	1.759
Mangrove Seagrass	1.759
Coral Coastal Forest	1.638
Coral Low Dune	1.638
Mangrove Seagrass Coastal Forest	1.598
Coral Seagrass Coastal Forest	1.598
Coral Seagrass Low Dune	1.598
Coral High Dune	1.446
Coral Marsh	1.446
Coral Seagrass High Dune	1.409
Coral Seagrass Marsh	1.409
Coral Seagrass Low Dune Coastal Forest	1.446
Mangrove Coral	1.194
Coral Seagrass High Dune Coastal Forest	1.264
Mangrove Coral Seagrass	1.160
Mangrove Coral Seagrass Coastal Forest	1.025

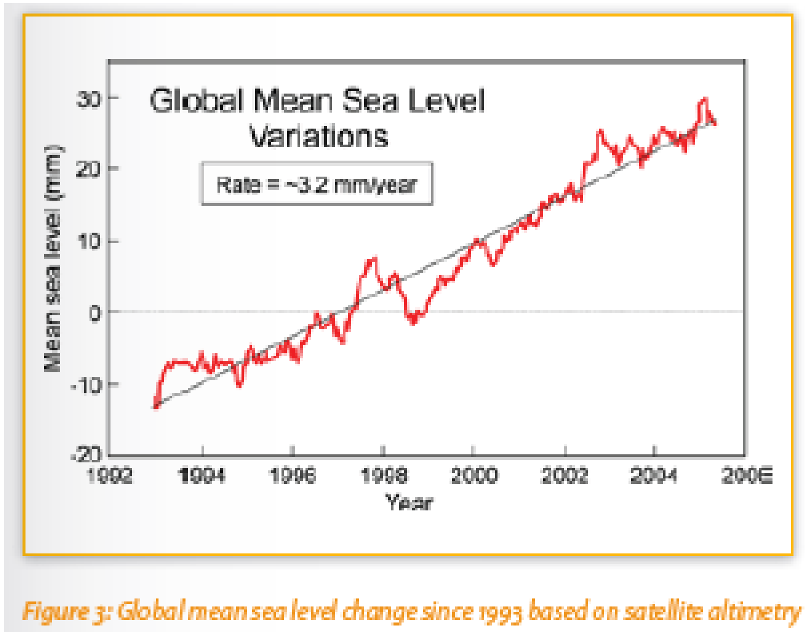


If users decide to create a similar layer, it is recommend that they create it in Microsoft Excel, and add the sheet in the "Layer" menu. To plot the data, right-click on the sheet name, and choose "Display XY Data". Choose to display the X and Y fields as "LONG" and "LAT", respectively. If users are satisfied with the result, right-click on the layer, choose "Export Data" and convert this temporary "Events Layer" into a point shapefile that can now be called when running the Coastal Vulnerability model. Finally, we recommend to use a WGS84 datum.

As described in *The Model* section cv-winds, the model provides an optional map of areas that are exposed or sheltered. This is purely based on fetch distances, and does not take into account measurements of wind speeds. To prepare this map, the model uses an estimate of a fetch distance cutoff to use that the user has defined, based on the AOI under consideration. To provide that distance, it is recommend that the "distance tool" on the global polygon layer, zoomed into the AOI, is used to determine that distance.

Sea level change

As mentioned earlier, a map of net rates of sea level rise or decrease in the AOI can be added. Such information can be found in reports or publications on Sea Level Change or Sea Level Rise in the region of interest. Otherwise, it is suggested that users generate such information from tide gage measurements, or based on values obtained for nearby regions that are assumed to behave in a similar way. A good global source of data for tide gauge measurements to be used in the context of sea level rise is the Permanent Service for Sea Level. This site has corrected, and sometimes uncorrected, data on sea-level variation for many locations around the world. From the tide gage measurements provided by this website, it is suggested that users estimate the rate of sea level variation by fitting these observations to a linear regression, as shown in the figure below. This figure was extracted from Bornhold (2008).



Create a sea level change GIS layer

Users can create their own polygon or points to represent the sea level change input to the model. If a polygon feature class is created, the model will apply the sea level change ranking assigned to that polygon for the segments of the shoreline that the polygon overlaps. If a point feature class is used, the model will assign sea level change rankings to shoreline segments based on whichever point is closest to that segment. To create a feature class in ArcMap, the map window must be in "data view" mode. Select the "Drawing" drop-down option and begin creating a polygon similar to the black feature below. Double click to complete the polygon. Similarly, you can select to create a 93marker94 rather than a 93polygon94 in the drawing tool bar. Select this option and click in locations throughout your area of interest where you would like to assign sea level change values or rankings. Next, click "Drawing >> Convert Graphics to Features..." Specify the path of the output shapefile or feature class and a name that will clearly designate the extent. Finally, check the box: "Automatically delete graphics after conversion" and click "OK". Once all polygons or points for specific regions are created, you must create an attribute field called "Trend" and populate it with values indicating net sea level change in mm/year according to *Table 4.1*. For general help with creating and editing shapefiles, try documentation provided by ArcGIS or QGIS. For more specific InVEST-related GIS video tutorials, consider enrolling in the online course, Introduction to the Natural Capital Project Approach.

Surge potential

Surge potential is estimated as the distance between a shoreline segment and the edge of the continental shelf, or any other depth contour of interest. This output is computed using a method that does not take into account the presence of land barriers between a shoreline segment and the depth contour.

When creating an AOI, loading the global polygon layer and the continental shelf (or other preferred depth contour, input 11) as guides is recommend. Draw the AOI so that it overlaps the portion of coastline you want to include in your analysis. Additionally, if you want to include the surge potential variable make sure the AOI overlaps at least a portion of the shelf's closest edge to the coastline. This is necessary so that the model can properly calculate the distance to shelf.

6.2.10 References

Bornhold, B.D., 2008, Projected sea level changes for British Columbia in the 21st century, report for the BC Ministry of Environment.

Bush, D.M.; Neal, W.J.; Young, R.S., and Pilkey, O.H. (1999). Utilization of geoindicators for rapid assessment of coastal-hazard risk and mitigation. Oc. and Coast. Manag., 42.

Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro Internacional de Agricultura Tropical (CIAT) (2005). Gridded Population of the World Version 3 (GPWv3). Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University.

Cooper J., and McLaughlin S. (1998). Contemporary multidisciplinary approaches to coastal classification and environmental risk analysis. J. Coastal Res. 14(2):512-524

Gornitz, V. (1990). Vulnerability of the east coast, U.S.A. to future sea level rise. JCR, 9.

Gornitz, V. M., Beaty, T.W., and R.C. Daniels (1997). A coastal hazards database for the U.S. West Coast. ORNL/CDIAC-81, NDP-043C: Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Hammar-Klose and Thieler, E.R. (2001). Coastal Vulnerability to Sea-Level Rise: A Preliminary Database for the U.S. Atlantic, Pacific, and Gulf of Mexico Coasts. U.S. Geological Survey, Digital Data Series DDS-68, 1 CD-ROM

Irish, J.L., and Resio, D.T., "A hydrodynamics-based surge scale for hurricanes," Ocean Eng., Vol. 37(1), 69-81, 2010.

Keddy, P. A. (1982). Quantifying within-lake gradients of wave energy: Interrelationships of wave energy, substrate particle size, and shoreline plants in Axe Lake, Ontario. Aquatic Botany 14, 41-58.

Short AD, Hesp PA (1982). Wave, beach and dune interactions in south eastern Australia. Mar Geol 48:259-284

Tolman, H.L. (2009). User manual and system documentation of WAVEWATCH III version 3.14, Technical Note, U. S. Department of Commerce Nat. Oceanic and Atmosph. Admin., Nat. Weather Service, Nat. Centers for Environmental Pred., Camp Springs, MD.

U.S. Army Corps of Engineers (USACE). 2002. U.S. Army Corps of Engineers Coastal Engineering Manual (CEM) EM 1110-2-1100 Vicksburg, Mississippi.

Wessel, P., and W. H. F. Smith (1996). A Global Self-consistent, Hierarchical, High-resolution Shoreline Database, J. Geophys. Res., 101, #B4, pp. 8741-8743.

World Resources Institute (WRI) (2009). "Value of Coral Reefs & Mangroves in the Caribbean, Economic Valuation Methodology V3.0".

6.3 InVEST GLOBIO Model

6.3.1 Summary

The GLOBIO model provides an index of biodiversity according to mean species abundance (MSA), the average population-level response across a range of species, to different stressors, including land-use change, fragmentation, and infrastructure. The model can be used as a static assessment of how far below a pristine state the current environment is or to estimate how a change in any of the stressors would lead to a stress in biodiversity or ecosystem integrity, as indicated by MSA.

6.3.2 Introduction

The GLOBIO methodology was developed by the United Nations Environmental Programme (UNEP, Alkemade et al. 2009) to model human impacts on biodiversity, measured by mean species abundance (MSA). Mean Species Abundance is an improvement over the more traditional species-area curve approach for two reasons. First, it gives aggregate estimates of species densities, not just species presence, which is important to represent true diversity since presence alone gives limited information about population viability (Balmford et al 2012). Second, it relates more than habitat area to changes in biodiversity by including information about the impact of fragmentation and threats from infrastructure (and climate change and nitrogen deposition if the full version of GLOBIO were implemented).

6.3.3 The model

The GLOBIO method consists of a set of equations linking environmental drivers to biodiversity impact, tables of parameters to estimate the above equations, based on a broad literature review, and suggested methodologies for inputting and processing the spatial data required. We have extended the GLOBIO methodology to downscale their global approach to a landscape level.

How it works

The GLOBIO approach is based on mean species abundance (MSA, see Schwartz et al. 2003 for an example usage of MSA). An MSA estimation ranges from 0 to 1, indicating the average proportional change in abundance of individual species in a location against that same location being pristine vegetation. An MSA of 1.0 implies that on average, species abundances are the same as in pristine land while an MSA of 0.0 implies that average species abundance is zero. Comparing MSA values of different scenarios identifies how anthropogenic changes affect biodiversity (e.g., changing from a scenario with a mean MSA of 0.6 to a new scenario with 0.5 implies that, on average, individual species' abundances declined 16.6% due to the land use change). In GLOBIO, an MSA value is defined for every grid-cell within a geospatial extent.

Stressors decrease MSA in a multiplicative way. In the GLOBIO3 paper, the stressors included land use/land cover (LU), excess atmospheric nitrogen deposition (N), proximity of infrastructure (mainly roads; I), fragmentation (F) and climate change (CC), as in the following equation to calculate MSA per pixel (i):

$$\mathrm{MS}A_i = MSA_{LU_i} \bullet MSA_{N_i} \bullet MSA_{I_i} \bullet MSA_{F_i} \bullet MSA_{CC_i}$$

To consider changes in land-use, we ignore the nitrogen deposition term and climate change terms; since neither of these terms change with each land-change scenario, they will cancel out when percent change in total (summed) MSA is calculated.

We refine the GLOBIO methodology for MSA change due to infrastructure, fragmentation, and land-use in order to make use of higher resolution land-use/land-cover data (500 m pixels from MODIS rather than 50 km pixels used by UNEP) needed to detect finer-scale ecological response that may include nonlinearities. Downscaling requires new methods for assigning land management regime sub-classes with more precision based on high-resolution data rather than continent-wide aggregates, and a more sophisticated approach for quantifying fragmentation than applying overall averages of patch size for different habitats.

Calculating MSA Impact from Infrastructure

Table 1 provides the data from Alkemade et al. (2009) for MSA values of different buffers around infrastructure that impact different ecosystems. The impact of infrastructure on MSA is determined solely by distance, not by the nature of the intervening vegetation. An area of cropland that is 5 km from a road will have its MSA reduced by a factor of 0.9 regardless of whether the area between the cropland and the road is tropical forest or more cropland. All sources of infrastructure are aggregated into a "man-made" land use/land cover class. The remaining land cover classes, which can then be considered vegetative or "natural" are split into three basic types: tropical forest, temperate or boreal forest, and grassland or cropland. The distance of these different habitat classes from infrastructure is used to calculate the impact zone for determining MSA from infrastructure, using Table 1.

Table 1: Effect of infrastructure impact zones on MSA, source: Alkemade et al. (2009)

Impact Zone	Tropical Forest Distance to infrastructure (m)	Temperate Forest Distance to infrastructure (m)	Grassland & Cropland Distance to infrastructure (m)	MSA_I	Standard Error
High Impact	<1000	<300	<500	0.4	0.22
Medium Impact	1000-4000	300-1200	500-2000	0.8	0.13
Low Impact	4000-14,000	1200-4200	2000-7000	0.9	0.06
No Impact	>14,000	>4200	>7000	1.0	0.02

Calculating MSA Impact from Fragmentation

We augment the standard GLOBIO approach to fragmentation analysis by using a fragmented forest quality index (FFQI). The FFQI is similar to methods used in the forestry literature, and is calculated by considering how many of a forest's neighboring cells are also forested. Rather than identifying the expected MSA impact from patch-size (as in GLOBIO), the FFQI estimates the relative effect of fragmentation with a Gaussian smoothing function. This treats habitat patches that are separated by only very small patches of infrastructure or non-habitat as less fragmented that habitat patches separated by wider distances. We convert the FFQI values on our map to km^2 to match the zones defined in Table 2 (according to Alkemade 2009) by taking the square root of the area to convert it to the width/height of the patch. Although the method is different from how UNEP defined patches, comparisons to the literature showed FFQI to be an accurate approximation of more cumbersome patch-based approach.

Table 2: Fragmentation effect on MSA under varying patch sizes, source: Alkemade et al. (2009)

FFQI	Area (km^2)	MSA_F	Standard Error
< 0.43	< 1	0.3	0.15
0.43 – 0.58	<10	0.6	0.19
0.58 – 0.90	<100	0.7	0.19
0.90 – 0.98	<1,000	0.9	0.20
0.98 – 0.99	<10,000	0.95	0.20
0.99 – 1	>10,000	1.0	0.20

Calculating MSA Impact from Land Use Change

The most difficult aspect of GLOBIO to implement is assigning different land-use/land-cover categories that relate to intensity of management or human use, since this information is often absent in remotely-sensed global land-cover datasets. To assist with this classification, we developed simple rules for reclassifying the MODIS or other satellite land-use/land-cover maps into the management categories for which MSA is quantified by GLOBIO's broad literature reviews. Table 3 presents the rule-based categorization used to convert MODIS data to GLOBIO-compatible classes. LULC types that are mapped to more than one GLOBIO type are then split according to other auxiliary datasets described below.

Forests: To distinguish between primary forest and other forest, including secondary (replanted) forests or forests with some extractive use and plantation forests, we analyze fragmentation in forest cover using FFQI and assign different use categories based on FFQI, with primary forest above a certain user-defined threshold. This approach assumes that pristine forests are more likely to be found in large, unfragmented tracts of forest, and that secondary or lightly used forests are more likely to be found in the most highly fragmented patches of forest. The threshold can be calibrated such that the aggregate amount of primary and secondary or lightly-used forests match estimates at the national or continental scale (documented in Alkemade et al. 2009).

Shrubland and Grassland: To distinguish between primary vegetation (more pristine) grasslands, grazed grasslands, and man-made pastures (deforested areas used for pasture), we compare the potential vegetation map generated by Ramankutty and Foley (1999) described above to actual vegetation determined by MODIS land-cover data. If a particular pixel is designated forest according to the potential vegetation map, but is listed as grassland in MODIS, it has likely forest that has been cleared for grazing, in this case the pixel is reclassified as "man-made pasture." If a pixel is grassland according to the potential vegetation map and is listed as grassland in the MODIS data, a separate dataset is utilized, quantifying the proportional pasture area at ~10 km resolution developed by Ramankutty et al. (2008). This pixel is defined as "livestock grazing" if the proportion of the grid-cell in pasture is greater than a user-defined threshold. The threshold can be chosen such that aggregate totals of livestock grazing match national and provincial data, as described above for forests. If the grassland pixel is lower than the grazing threshold, it will be defined as primary vegetation.

Cropland: Because cropland intensification is only calculated in the MSA_{LU} and does not affect the configuration of primary habitat and thus the fragmentation calculated for MSA_F, the spatial location of intensification is not necessary to define. The user only needs to designate the proportion of agriculture in the landscape that is intensified (i.e., not low-input agriculture). This can be found in the regional datasets cited by Alkemade et al. (2009) or available through FAO, or can be derived a dataset developed by Foley et al. (2011) that maps yield gaps for all major commodity crops globally at ~10 km resolution. This methodology compares agricultural production in similar climates (based on precipitation and growing degree days) and rates crop yield in different regions according to the maximum yields attained for its particular climate. The difference between actual and maximum attainable yield is defined as the "yield gap." The yield gap can serve as a surrogate for (lack of) intensification, and the user can examine the yield gap maps for their region of interest to determine what proportion of the landscape falls below a certain level of yield gap.

Table 3: MODIS to GLOBIO cover class conversion and MSA affected by land use

MODIS Land Use/Land Cover Class	Convert to which GLOBIO classes?	MSA_LU	SE
0 - Water	N/A		
1 - Evergreen needleleaf forest 2 - Evergreen broadleaf forest 3 - Deciduous needleleaf forest 4 - Deciduous broadleaf forest 5 - Mixed forest	1 - Primary vegetation[a] 3 - Secondary forest[a]	1 0.5 0.2	<0.01 0.03 0.04
6 - Closed shrublands/cerrado 7 - Open shrublands 8 - Woody savannas 9 - Savannas 10 - Grasslands	1 - Primary vegetation[b] 5 - Livestock grazing[c] 6 - Man-made pastures[b]	1 0.7 0.1	<0.01 0.05 0.07
12 - Croplands/Perennial	12 – All agriculture	0.3 0.1	0.12 0.08
13 - Urban and built-up	10 - Built-up areas	0.05	
16 - Barren or sparsely vegetated	1 - Primary vegetation	1.0	<0.01

Split based on (a)FFQ (described in Fragmentation section, above), (b) potential vegetation map (Foley et al. 2009), (c) proportional pasture area (Ramunkutty et al. 2009). Missing from this classification structure is GLOBIO classes "Lightly used natural forest" (GLOBIO class 2), "Plantation forest" (GLOBIO class 4), and "agroforestry" (GLOBIO class 7), "Low-input agriculture" (GLOBIO class 8), and "Intensive agriculture" (GLOBIO class 9). The agriculture classes are split in an aspatial calculation of MSA_LU according to the "Fraction of intensification" value set by the user.

Limitations and simplifications

MSA is an aggregate estimate, making it impossible to track compositional effects, and there are many different compositional possibilities for the same MSA. While MSA caps relative abundance of individual species at 1, ensuring that a local rise in one species cannot disguise a fall in overall species abundance, an MSA of 0.5 could mean that all species are half as abundant as in a pristine state, or that one species has suffered immense decline while the rest have remained constant, or anywhere in between. Additional information about the shape of the distribution of species abundances and extinction probabilities related to different levels of MSA could improve the usefulness of this index. But even then, diversity is more complex than numbers of species and population numbers. Some conservation biologists argue that species composition is as important as any other measure of diversity, and tracking specific species is essential to estimating impacts on threatened or endangered species or culturally valuable species (Phalan et al. 2011 Food Policy). To achieve this level of specificity, the impacts of different land-use strategies would need to be evaluated for each species individually and then combined across species for summary results, which may not be possible in many regions of the world with low data availability and high agricultural and other development pressure. In such cases, MSA provides a quick and easy to use index for biodiversity change in decision contexts.

In our application of GLOBIO, we use the mean parameter values and their standard errors to estimate the impacts of infrastructure, land-use, and fragmentation at new locations, which assumes that these values represent a random sample of species and geographic locations. However, limited data availability for certain taxonomic groups and geographic regions mean that there are potential biases in the parameter estimates that add an unquantifiable degree of uncertainty to predictions based on our application of GLOBIO.

The estimates of the impact of infrastructure are based on a meta-analysis of ~75 studies, predominately of bird and mammal populations in Europe and North America, with some information on insects and plants (Alkemade et al. 2009; Benítez-López et al 2010). Whether the impacts of infrastructure are similar for other taxonomic groups or geographic areas is unknown.

Estimates of the impacts of land use are based on a slightly greater number of studies, with 89 identified in the initial publication of GLOBIO (Alkemade et al. 2009) and 195 identified in a final published meta-analysis (de Baan et al. 2013). The parameter estimate for all artificial surfaces/built-up areas was based on expert opinion, representing densely populated cities, and without quantification of uncertainty (Alkemade et al. 2009). Datasets come largely from tropical regions, with fewer from temperate regions and none from boreal zones (de Baan et al. 2013). Data were available for 9 out of 14 biomes, and for many biomes, information was only available for some land use types. For example, information on permanent crops, agroforestry and artificial areas came only from two biomes. For three biomes, information was only available for pastures, but not for other land use types. As is common, data were also taxonomically biased towards vertebrate and plant species (de Baan et al. 2013). Arthropods were under-represented, and bacteria and fungi were not included at all in the database.

Furthermore, our assignment of satellite land-cover (e.g., forest or grassland) to the different GLOBIO land-use classes (e.g., primary vs. secondary forest or pristine vs. grazed grassland) introduce additional error that is not incorporated into the analysis. While we can ensure that our assignments aggregate up to national or regional level statistics, we cannot ground-truth our classification system to quantify the level of accuracy or uncertainty.

The impacts of fragmentation on mean species abundance (MSA) are based on six datasets from 3 publications. The proportion of species with a viable population was used as a proxy for MSA (Alkemade et al. 2009), and it is unclear how much additional uncertainty in the parameters that adds. Taxonomic and geographic biases are again a limitation. Two studies focus exclusively on mammals, including ~30 mammal species in Florida (Allen et al. 2001) and 10 species of carnivores from around the world (Woodroffe & Ginsberg 1998). The third study is limited exclusively to Europe, of which half of the 202 species included are birds (Bouwma et al. 2002).

6.3.4 Data needs

The model uses 11 forms of input data. 3 are required and 8 are optional. **NOTE: All spatial data must be projected in meters (i.e., a local, not a global or lat-long projection), to ensure accurate distance to infrastructure calculations. The model will not execute without a defined projection.**

1. Land-use/cover map (required), following one of two options:

   (a) Vegetation-specific (not management-specific) land-cover. This is the type of land-cover you may acquire from MODIS or other remotely-sensed data sources. It distinguishes between forest, grassland, savanna, cropland, and other vegetation types. It does NOT distinguish between the differences in management defined by GLOBIO, such as primary vs. secondary vegetation, or grassland vs. pasture. If this option is chosen, several helper datasets (listed as required for option 1a, below) will be required.

(b) Management-specific land-cover, following the classification scheme established by GLOBIO (see Table 3, above). If this option is chosen, tick the box for "Predefined land use map for GLOBIO" and enter the map there. All other data inputs will turn grey except for the other required data set, the infrastructure directory, and the optional AOI input.

Name: file can be named anything (lulc_2008.tif in the sample data)

Format: standard GIS raster file (e.g., ESRI GRID or IMG), with a column labeled 'value' that designates the LULC class code for each cell (integers only; e.g., 1 for forest, 10 for grassland, etc.) The LULC 'value' codes must either match the LULC class codes used in the Land-cover to GLOBIO land-cover table described below (if choosing option 1a) or the GLOBIO land-cover specified in Table 3 (if choosing 1b). The table can have additional fields, but the only field used in this analysis is one for LULC class code.

1. Infrastructure directory (required). This is a folder containing maps of any forms of infrastructure you wish to consider in the calculation of MSA_I. These data may be in either raster or vector format.

   Name: folder can be named anything (infrastructure_dir in the sample data)

   Format: the files within the folder can be either raster or vector

1. Land-cover to GLOBIO land-cover table (required for option 1a). This is a table that translates the land-cover of option (a) in (1) above to intermediate GLOBIO classes, from which they will be further differentiated using the additional data below.

   Name: file can be named anything (lulc_conversion_table.csv in the sample data)

   File type: *.csv

   Rows: each row is a different LULC class.

   Columns: the columns must be named as follows:

1. lucode: Land use and land cover class code of the dataset used. LULC codes match the 'values' column in the LULC raster of (1a) and must be numeric and unique.
2. globio_lucode: The LULC code corresponding to the GLOBIO class to which it should be converted, using intermediate codes described in the example below.

   Example: On the left is MODIS land-cover data, using the UMD classification, as defined in Table 3. On the right is the GLOBIO land-cover translation, which lumps the forest classes (1-5 in MODIS) into 130, grassland/shrubland (6-10 in MODIS) into 131, and agriculture (12 in MODIS) into 132. Urban land-use (13 in MODIS) maps directly onto built-up lands (10 in GLOBIO). Barren or sparsely vegetated (16 in MODIS) can be treated primary vegetation (1 in GLOBIO). The subsequent datasets and/or user inputs will help determine how to split up the 130, 131, and 132 into primary and secondary vegetation, rangelands and pasture, and intensified and unintensified agriculture, respectively.

lucode	globio_lucode
0	0
1	130
2	130
3	130
4	130
5	130
6	131
7	131
8	131
9	131
10	131
12	132
13	10
16	1

1. Pasture map (required for option 1a). The proportional pasture area, as developed by Ramankutty et al. (2008). See explanation in *Shrubland and grassland* under *How it Works*, above.

   Name: file can be named anything (pasture.tif in the sample data)

   Type: standard GIS raster file (e.g., ESRI GRID or IMG), with a column labeled 'value' that designates the proportion of the pixel that is in pasture (restricted to floats between 0 and 1).

2. Potential vegetation map (required for option 1a). This should be the potential vegetation map generated by Ramankutty and Foley (1999), or similar approach. It is important to use either this exact map or if using a different method for mapping potential vegetation, convert the land cover classifications to match those of this map. See explanation in *Shrubland and grassland* under *How it Works*, above.

   Name: file can be named anything (potential_vegetation.tif in the sample data)

   Type: standard GIS raster file (e.g., ESRI GRID or IMG), with a column labeled 'value' that designates the land cover class (integers only) according to Ramankutty and Foley (1999).

3. Primary Threshold (required for option 1a): a value between 0 and 1 that will determine the FFQI (forest fragmentation quality index) at which a cell should be assigned to primary or secondary forest, which can be adjusted such that the aggregate land-use matches regional statistics.

4. Pasture Threshold (required for option 1a): a value between 0 and 1 that will determine the proportion of pasture within a cell (in the pasture map, input #4) in order for that cell to be assigned to grassland or livestock grazing, which can be adjusted such that the aggregate land-use matches regional statistics.

5. Proportion of Agriculture Intensified (required for option 1a): a value between 0 and 1 denoting the proportion of total agriculture that should be classified as "Intensive agriculture" or GLOBIO class 8 (with 1 – Proportion of Agriculture Intensified being the proportion classified as "Low-input agriculture", GLOBIO class 9) in the computation of MSA_{LU}.

6. MSA parameter table (required). This table sets the values for MSA that should be used for the different impacts (infrastructure, fragmentation and land-use) to compute MSA_I, MSA_F, and MSA_{LU}. The example below (and in the sample data) gives the mean values and standard errors provided in Alkemade et al. (2009). This table can be altered to put high and low estimates from confidence intervals in the msa_x column, to aid in uncertainty assessment.

   Name: file can be named anything (msa_parameters.csv in sample data)

   Type: *.csv

   Columns: the columns must be named as follows:

(a) MSA_type: either msa_i_primary, msa_i_other, msa_f, or msa_lu. The values for msa_i are taken from Table 1 above, and msa_i_primary in the example below corresponds to the values used for tropical forest and msa_i_other corresponds to values used for grassland and cropland.

(b) Measurement: the metric by which the value in the subsequent column is measured.

(c) Value: the level of impact from which the MSA value is derived (e.g., m of distance from infrastructure for msa_i, the FFQI

(d) MSA_x: the MSA set by Alkemade et al. (2009) for different types of impacts

(e) SE: the standard error associated with each MSA value, according to the meta-analysis in Alkemade et al. (2009). These values are not used by the model but are recorded here in this sample data set so that the user can adjust the MSA_x values according to the confidence interval.

Example:

MSA_type	Measurement	Value	MSA_x	SE
msa_i_primary	Distance (m)	<1000	0.4	0.22
msa_i_primary	Distance (m)	1000-4000	0.8	0.13
msa_i_primary	Distance (m)	4000-14000	0.9	0.06
msa_i_primary	Distance (m)	>14000	1	0.02
msa_i_other	Distance (m)	<500	0.4	0.22
msa_i_other	Distance (m)	500-2000	0.8	0.13
msa_i_other	Distance (m)	2000-7000	0.9	0.06
msa_i_other	Distance (m)	>7000	1	0.02
msa_f	FFQI	< 0.43	0.3	0.15
msa_f	FFQI	0.43 - 0.58	0.6	0.19
msa_f	FFQI	0.58 - 0.90	0.7	0.19
msa_f	FFQI	0.90 - 0.98	0.9	0.2
msa_f	FFQI	0.98 - 0.99	0.95	0.2
msa_f	FFQI	0.99 - 1	1	0.2
msa_lu	Land Cover Class	0	0	
msa_lu	Land Cover Class	1	1	<0.01
msa_lu	Land Cover Class	2	0.7	0.07
msa_lu	Land Cover Class	3	0.5	0.03
msa_lu	Land Cover Class	4	0.2	0.04
msa_lu	Land Cover Class	5	0.7	0.05
msa_lu	Land Cover Class	6	0.1	0.07
msa_lu	Land Cover Class	7	0.5	0.06
msa_lu	Land Cover Class	8	0.3	0.12
msa_lu	Land Cover Class	9	0.1	0.08
msa_lu	Land Cover Class	10	0.05	na

1. AOI – Area of Interest (optional). If a summary of the MSA value is desired for the region, click the box next to AOI and enter a vector dataset containing the area(s) of interest, either as a region area or partitioned into subregions (e.g., ecoregions, districts, etc.).

   Name: file can be named anything (sub_aoi.shp in the sample data)

   Type: polygon (vector) data

6.3.5 Running the model

The model is available as a standalone application accessible from the install directory of InVEST (under the subdirectory invest-3_x86/invest_globio.exe).

Viewing Output from the Model

Upon successful completion of the model, a file explorer window will open to the output workspace specified in the model run. This directory contains an output folder holding files generated by this model. Those files can be viewed in any GIS tool such as ArcGIS, or QGIS. These files are described below in Section Interpreting Results.

6.3.6 Interpreting Results

Final Results

Final results are found within the *Workspace* specified for this module.

- **globio-log: Each time the model is run, a text (.txt) file will** appear in the *Output* folder. The file will list the parameter values for that run and will be named according to the service, the date and time, and the suffix.
- **aoi_summary_<suffix>: A shapefile summarizing the average MSA** for each zone defined in the area of interest.
- **msa_<suffix>.tif: A raster of the overall MSA (mean species** abundance) value, defined as "the average abundances of originally occurring species relative to their abundance in the original, pristine or mature state as the basis." This index is on a scale of 0 to 1, with 1 being the pristine condition, calculated as the product of the MSA_{LU}, MSA_F, and MSA_I below.
- **msa_lu_<suffix>.tif: A raster of MSA calculated for impacts of** land-use only.
- **msa_f_<suffix>.tif: A raster of MSA calculated for impacts of** fragmentation only.
- **msa_i_<suffix>.tif: A raster of MSA calculated for impacts of** infrastructure only.

Intermediate Results

You may also want to examine the intermediate results. These files can help determine the reasons for the patterns in the final results. They are found in the *intermediate_outputs* folder within the *Workspace* specified for this module.

- **combined_infrastructure_<suffix>.tif: A map joining all the** infrastructure files in the infrastructure directory (input 2 above). If there is only one file in that directory, it should be identical to that file.
- **distance_to_infrastructure_<suffix>.tif: A map coding each** pixel by its distance to the nearest infrastructure, used to compute MSA_I. Distance in this raster is measured as number of pixels, which is converted to meters in the model using the defined projection.
- **ffqi_<suffix>.tif: A map of the forest fragmentation quality** index (ffqi), used to differentiate between primary and secondary forest in the GLOBIO land use classification.
- **globio_lulc_<suffix>.tif: The final land use map converted to** GLOBIO classification, as outlined in Table 3. If desired, this map (or any altered version of this map) could be used to run the model using option 1b, above. This is used to compute MSA_{LU}.
- **primary_veg_smooth_<suffix>.tif: A Gaussian-filtered** ("smoothed") map of primary vegetation (identified in globio_lulc), used to compute MSA_F.

6.3.7 References

Alkemade, Rob, Mark van Oorschot, Lera Miles, Christian Nellemann, Michel Bakkenes, and Ben ten Brink. "GLOBIO3: a framework to investigate options for reducing global terrestrial biodiversity loss." *Ecosystems* 12, no. 3 (2009): 374-390.

Allen, C. R., Pearlstine, L. G., & Kitchens, W. M. (2001). Modeling viable mammal populations in gap analyses. Biological Conservation, 99(2), 135–144. doi:10.1016/S0006-3207(00)00084-7

Balmford A., R. Green, B. Phalan. 2012 What conservationists need to know about farming. Proc. R. Soc. B 279: 2714–2724.

Benítez-López, A., Alkemade, R., & Verweij, P. a. (2010). The impacts of roads and other infrastructure on mammal and bird populations: A meta-analysis. Biological Conservation, 143(6), 1307–1316. doi:10.1016/j.biocon.2010.02.009

Bouwma, I. M., Jongman, R. H. G., & Butovsky, R. O. (2002). Indicative map of the Pan-European Ecological Network - technical background document. Tilburg, The Netherlands/Budapest, Hungary.

de Baan, L., Alkemade, R., & Koellner, T. (2013). Land use impacts on biodiversity in LCA: a global approach. International Journal of Life Cycle Assessment, 18, 1216–1230. doi:10.1007/s11367-012-0412-0

Foley , J.A., et al. 2005. Global consequences of land use. Science 305: 570-574.

Foley, J.A., et al. 2011. Solutions for a cultivated planet. Nature 478: 337-342.

Mueller, N., et al. 2012. Closing yield gaps through nutrient and water management. Nature 490: 254-257.

Phalan, B., A. Balmford, R.E. Green, J.P.W. Scharlemann. 2011. Minimising the harm to biodiversity of producing more food globally. Food Policy 36: S62-S71.

Ramankutty, N. and J.A. Foley. 1999. Estimating Historical Changes in Global Land Cover: Croplands from 1700 to 1992, Global Biogeochemical Cycles, 13 (4), 997-1027

Ramankutty, N., et al. 2008. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. Global Biogeochemical Cycles, Vol. 22, GB1003

Woodroffe, R., & Ginsberg, J. R. (1998). Edge Effects and the Extinction of Populations Inside Protected Areas. Science, 280(5372), 2126–2128. doi:10.1126/science.280.5372.2126

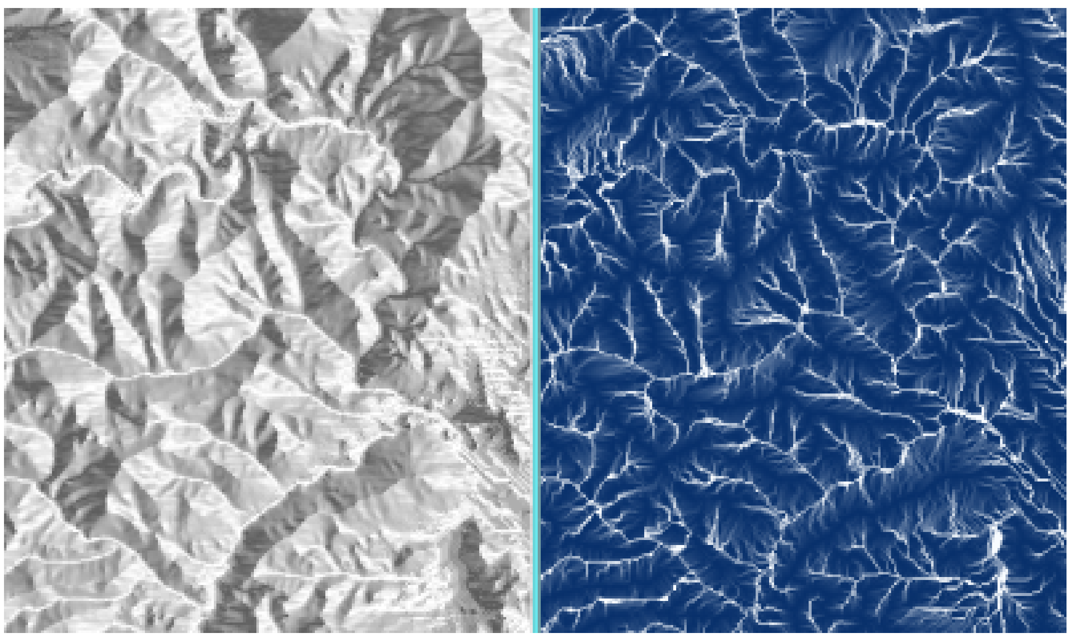
CHAPTER

SEVEN

SUPPORTING TOOLS

7.1 RouteDEM

7.1.1 Summary



7.1.2 Introduction

The freshwater models in InVEST are routed across a DEM. In the course of developing InVEST we've found existing implementations of flow direction and flow accumulation algorithms to be lacking. To address this need, we have developed our own high performance implementations of the d-infinity flow direction algorithm (Tarboton 1997), combined with a plateau resolution algorithm to route across flat areas (Garbrecht and Martz) that outperforms TauDEM and GRASS implementations. We feel these implementations are useful enough that we offer them as a standalone tool that can calculate:

- Resolving flat areas in a DEM so that all regions, except pits, drain to some point.
- The d-infinity flow direction across an arbitrary DEM.
- The d-infinity flow accumulation algorithm across a DEM.

- Stream thresholding of the flow accumulation algorithm.

After installing InVEST, this tool can be found in the start menu under the InVEST folder as a utility called RouteDEM.

7.1.3 Tool Inputs

1. **Workspace**: This is the folder that will contain outputs from RouteDEM after it is run.
2. **DEM**: A GIS DEM raster input. For a good route, the DEM should first be pit filled. Flat plateau regions will be automatically resolved by RouteDEM.
3. **Plateau Resolved DEM Filename**: This is the name of the output file that the plateau resolved dem will be saved to in the workspace.
4. **Flow Direction Filename**: This is the name of the output file that the d-infinity flow direction raster will be saved to in the workspace.
5. **Flow Accumulation Filename**: This is the name of the output file that the d-infinity flow accumulation raster will be saved to in the workspace.
6. **Threshold flow accumulation**: This is the value that will be used to threshold the flow accumulation raster to create a stream layer. The output will be called stream_[threshold].tif in the workspace where [threshold] will be replaced by the value in this input.
7. **Calculate multiple stream thresholds**: If checked, multiple stream threshold rasters will be generated where

   7a. **Threshold Flow Accumulation Upper Limit** is the upper value of the multiple stream threshold set.

   7b. **Step size** is the number of threshold steps to take between rasters. If the original flow accumulation raster is 1000, the threshold upper limit is 2000 and the step size is 100, RouteDEM will generate 10 stream threshold rasters of limits 1000, 1100, 1200, ..., 2000. This can be useful to explore the space of this parameter for other InVEST inputs that require a threshold stream layer to be defined.
8. **Calculate slope**: If checked RouteDEM will also calculate the slope of the input DEM. If selected also define **Slope Filename** as the name of the output slope raster to be placed in the workspace.

7.1.4 References

Garbrecht, J., Martz, L. W.,The assignment of drainage direction over flat surfaces in raster digital elevation models, Journal of Hydrology, Volume 193, Issues 1–4, 1 June 1997, Pages 204-213, ISSN 0022-1694, http://dx.doi.org/10.1016/S0022-1694(96)03138-1. (http://www.sciencedirect.com/science/article/pii/S0022169496031381)

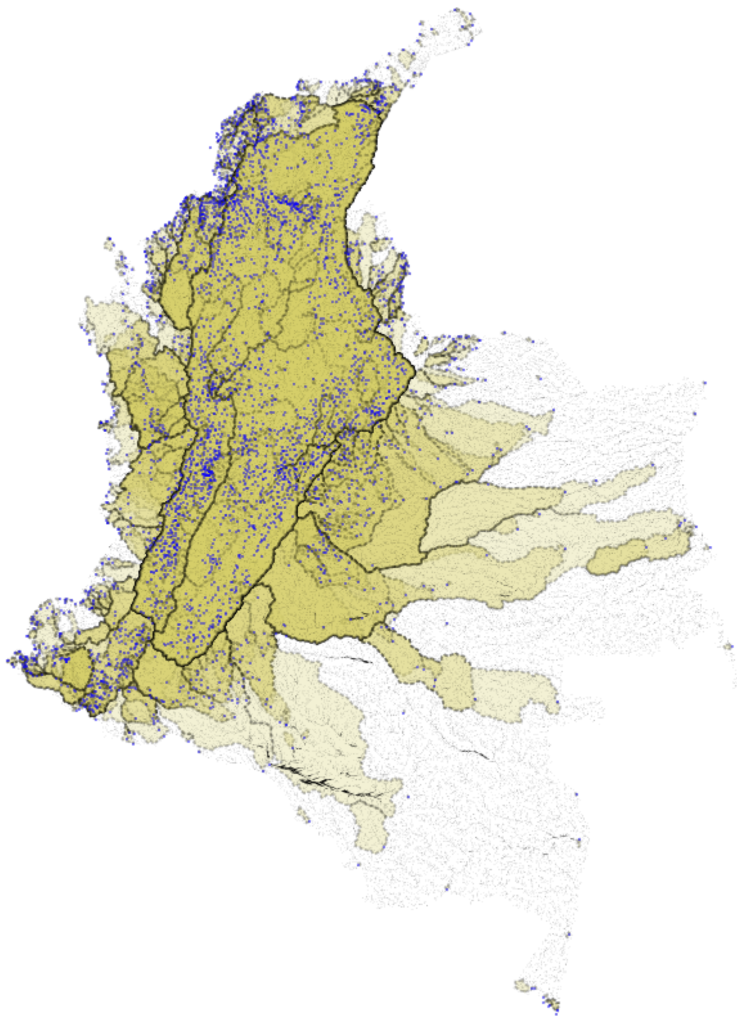
Tarboton, D. G. (1997), A new method for the determination of flow directions and upslope areas in grid digital elevation models, Water Resour. Res., 33(2), 309–319, doi:10.1029/96WR03137.

7.2 DelineateIT

7.2.1 Summary

7.2.2 Introduction

Many of the freshwater models in InVEST require watershed polygons to aggregate the ecosystem service provides to beneficiaries. However, we've found the creation of watersheds with existing tools to be difficult and often requiring specific expertise and/or proprietary toolsets. To address this need, we have developed our own watershed delination



algorithm released in the PyGeoprocessing GIS package, and wrapped into a UI inside of InVEST. All DEM routing is handled by PyGeoprocessing which resolves plateaus and uses d-infinity to route flow directions.

After installing InVEST, this tool can be found in the start menu under the InVEST folder as a utility called DelinateIT.

7.2.3 Tool Inputs

1. **Workspace**: This is the folder that will contain outputs from RouteDEM after it is run.
2. **DEM**: A GIS DEM raster input. The DEM should first be pit filled. Flat plateau regions will be automatically resolved by PyGeoprocessing.
3. **Outlet Points**: This is a point shapefile that is used to specify areas which the watersheds should be constructed around. These may be stream intake points, population centers, or other points of interest. Any fields associated with this dataset will be copied to the watershed shapefile as it is constructed.
4. **Threshold flow accumulation**: DelineateIT can snap outlet points to the nearest stream to make more robust watersheds than those that are off center of a pixel. To do this, the tool constructs a stream layer is constructed by thresholding any pixels whose flow accumulation values exceed this parameter.
5. **Pixel Distance to Snap Outlet Points**: During watershed construction, DelineateIT will search a window of this size to find the nearest stream pixel classified by the threshold flow accumulation given previously.

7.2.4 Tool Outputs

All outputs can be found in the **Workspace** directory given previously. They include:

- **watersheds.shp** this is a polygon shapefile defining the areas that are upstream from the snapped outlet points; upstream defined by the d-infinity flow algorithm implementation in PyGeoprocessing.
- **snapped_outlets.shp** this is a point shapefile that indicates where the outlet points were snapped to based on the values of "threshold flow accumulation" and "pixel distance to snap".
- **stream.tif** this is the raster stream layer used to snap outlet points.

7.2.5 References

PyGeoprocessing https://bitbucket.org/richpsharp/pygeoprocessing .. primerend

7.3 Scenario Generator

7.3.1 Summary

7.3.2 Introduction

Scenarios are storylines that depict future events and states. They provide insight into the future and help shape imagination about uncertain future events. For scenarios to have the desired effects and applicability, they must be plausible to let the audience relate to the possibility of such an event occurring; otherwise, they run the risk of being dismissed. As described in the InVEST scenario guidelines, scenarios describe a possible future, reflect important and uncertain future developments or choices, are plausible, internally consistent, and relevant to the questions being addressed. Scenarios are used in many fields and take different forms, but in this document, we consider scenarios that can be expressed in a spatially explicit manner and that relates to land cover.

Land use and changes in it is an issue of global importance. Land use/Land cover change is driven by the need to provide food and shelter and the rise in economic development among many other factors that change quickly and grow complex day by day. Today's action can drive land cover change in the future, and scenarios are important for raising awareness on possible consequences. In the context of ecosystem services, it is important to compare potential change in the provision of ecosystem services under various scenarios, hence to evaluate the impact of these alternatives and alert decision makers on the consequences of loss.

While land use change modeling is very important, it is complex and requires incorporation of social, political and economic factors that drive the change in land use/land cover. However, incorporating such factors is hampered by methodological difficulties and lack of spatially explicit data (Veldkamp and Lambin 2001). Due to this complexity, proxy variables, which are easier to measure (e.g. distance to roads), are used in place of the actual drivers. This approach runs the risk of obscuring causality, but provides simple direct pathways into understanding such a complex issue.

This scenario generation tool provides a relatively simple method of generating scenarios based on land suitability. It works on the principle that changes on land occur on areas that are relatively more suitable. By combining stakeholder input of transition likelihood with physical factors that determine suitability, The InVEST Scenario Generator generates simple maps that depict future land cover. At the heart of it is land suitability analysis.

Land suitability analysis has developed over the years starting from the early hand drawn maps advancing to computer assisted overlay mapping, multi criteria evaluation and currently artificial intelligence methods (Collins et. al. 2001). With modern developments in Geographic Information Systems (GIS), overlay analysis is commonplace and possible with almost every GIS software. One of the main limitations of classical overlay mapping and modeling approaches is the difficulty in incorporating value judgments such as decision maker preferences and underlying policy driver influence (Malczewski 2004). Combining GIS with multi-criteria evaluation – a method of evaluating various criteria in decision-making makes it easier to incorporate expert knowledge into such analysis. Others have favored artificial intelligence (AI) methods which mimic human intelligence without claiming an understanding of the underlying processes (Malczewski 2004). Such techniques include fuzzy logic, neural networks, genetic algorithms and cellular automata. These methods have been criticized for their black box nature of analysis, making them less easily accepted by decision makers. The Scenario Generator uses a combination of overlay analysis, multi-criteria evaluation methods and direct application of expert knowledge to map alternative futures.

7.3.3 The Model

How it Works

Allocating land parcels to various uses is a multi-objective multi-criteria problem. Landcover change is influenced by a large number of drivers. Some of them are very clear and easy to observe (e.g. mineral exploration and extraction) while others are subtle and difficult to identify (e.g. change in societal values). As a participatory process, the Scenario Generator relies on drivers that stakeholders can easily identify and estimate impacts of. At the same time, there are many factors that influence suitability of land parcels for conversion. This may include factors or constraints such as slope, soil type, distance to roads, distance to markets, rainfall distribution, and access. Dealing with multiple objectives and multiple criteria presents a challenge that The InVEST Scenario Generator attempts to solve.

The InVEST Scenario Generator is designed to work with data from experts/stakeholders typically in a workshop setting. The major components of the input are i) the transition likelihood, ii) the physical and environmental factors that influence change and iii) the quantity of anticipated change under a given scenario. The tool works off of a single land cover map making it useful in data-poor situations.

The following are the details of the elements needed to run the tool.

Quantity of Change

The quantity of change is determined indirectly by the demand for land and is estimated by stakeholders. The change value provided here is used as the goal, and the tool converts all suitable pixels in order of suitability until this goal is met for the cover or until all the available pixels are converted. For cover types that do not already exist in the initial landcover, the change quantity should be entered in absolute area units (Hectares). While this guide does not give explicit directions on how to estimate this quantity, users can use a variety of methods based on the level of sophistication they desire. As an example, the workshop facilitator or modeler can run an analysis of past land cover change to provide a basis of the magnitudes of change that are plausible in the landscape. However, the final values used in the tool should be determined by stakeholders and could differ from such empirical data. If taking a purely empirical modeling approach then there are other tools that may be better suited.

Transition Likelihood

The likelihood that a given parcel is converted from one land cover type to another (transition likelihood) is defined by the stakeholders. This exercise is preferably done in a group setting. The stakeholders should select a few well-understood transitions and determine how likely they are in the period under consideration. For example, they may consider the likelihood of grassland being converted to agriculture and on a scale of 0 to 10, and assign a value of 8. When deciding the transition likelihood, the stakeholders should consider only the drivers which influence such a change. Constraints such as protection should not be considered here, as they are estimated separately.

Following is an example of the transition likelihood matrix:

			To		
		Cover A	Cover B	Cover C	Change
	Cover A	0	4	0	30%
From	Cover B	0	0	0	0
	Cover C	10	2	0	-10%

The absolute value of the numbers in the matrix above is not critical because they are used relatively. It is important though to have these in the range of 0~10 that is used when creating the suitability maps (using factors as discussed later). Reading the columns, cover A will grow and all of its growth will be from cover C; Cover B will grow due to a contribution from Cover A and a smaller contribution from Cover B. As currently written, in this case cover B will only grow into cover C after all the available parcels in cover A are exhausted.

Priority

The Scenario Generator requires that the stakeholders rank the land cover types to assign weight. When multiple objectives compete for a single land parcel (or pixels as unit used in this tool), the one with the higher weight or priority wins. Priority ranking the cover types is difficult, and an optional feature is provided which utilizes a pair-wise comparison matrix in an analytic hierarchy process (AHP), such that the stakeholders only compare two cover types at a time. Using AHP is optional.

As an example, in the table below, using the 9 point continuous scale (see at end of document), Cover C is *"extremely less important"* compared to Cover A (0.11 versus 1). However, cover C is *"strongly more important"* compared to Cover B. Once the table has been filled the tool uses eigenvectors to assign weights to each of the cover types. It is recommended to use the 9-point continuous scale (Saaty 1977). Alternatively, one can derive the weights using any other tools and enter them directly. The cover type that has the highest weight will have its goal achieved before moving to the next. It is important to note that when the priorities are calculated, the input table is not updated but the calculated values are used by the tools when doing conversion.

id	Name	Cover A	Cover B	Cover C	Priority
1	Cover A	1			
2	Cover B	0.5	1		
3	Cover C	0.1	5	1	

The matrix above is used to compute the suitability. There will be as many suitability layers as the number of cover types (objectives) being considered, with values closer to 10 showing pixels that would be converted first.

Factors

The transition likelihood values given in table 1 are based on expert opinion and policy drivers. However, certain physical and environmental factors also affect the pixel suitability for conversion, hence determining where on the landscape the land cover changes are likely to happen. Examples of such factors include distance from roads, soil types, distance from cities, elevation, slope, and aspect. The tool allows the user to provide these factors and their relationship with land suitability. The impact of these factors differ between objectives (cover types), therefore the user can enter more than one factor for each of the cover types, and apply one factor to multiple cover types. The tool uses relative weights to combine these factors and determine the areas most suitable for certain land cover. The current version accepts vector layers as factor input. Point and line factors are used to calculate suitability based on proximity. Polygon input should have a field indicating suitability level ranging from 0 (unsuitable) to 10 (extremely suitable). The effects or all factors are then combined based on the weights defined by user. Selection of factors is very critical to producing plausible scenario maps. Using factors in the tool is optional and it is advised to use a good dataset, otherwise the results can be inconsistent. Given the iterative nature of scenario development, users should run the tool multiple times with and without the factors to see the effects and adjust as necessary.

Proximity Suitability

Pixels close to a land cover type may be more likely to be converted to that cover type. For example, parcels close to agriculture, if suitable for agriculture may be most likely to be converted first. However, this may not be the case for all cover types, therefore the user has the option to mark a cover type as having proximity suitability. While this proximity may play an important role in improving suitability of parcels, the effect is capped in this tool, so that it does not drive the suitability significantly. To apply the effect of proximity, the distance of each cell to the cover being analyzed is computed and made to diminish up to the maximum distance entered by the user. The cells closest to the cover are given the highest value while those farther than the maximum distance given a value of 1. When combining with the rest of the suitability layers, the effect of the proximity distance is made to slightly improve suitability.

Constraints

Constraints are unique factors that prevent human-induced land cover change. An example of a constraint is a protected area. However, protected areas have different designations which determine their ability to prevent land cover change. Even where they are gazetted as “strictly protected”, implementation on the ground may vary. Therefore the tool allows the user to enter an access value that determines the extent to which the protected area would effectively prevent habitat conversion under the scenario in consideration. An access value of 0 implies that the constraint has full effect and no conversion can take place within the boundary of the constraint while a value of 1 implies that the constraint has no effect. When applied to a suitability layer, a constraint of 0 makes all the parcels with which it overlaps to have a suitability of 0 thereby not having any chance of attracting any changes. An example of another type of constraint, albeit more complex is a requirement that only parcels beyond a specific area can be converted to large-scale agriculture. If such a constraint is applied, any suitable regions (group of pixels) that do not meet the minimum requirement are ignored.

Overview of the suitability components

For each target LULC type, the suitability of potential land parcels is calculated by:

$$S = \min(100, \{[C^*(L^* WL + F * WF) + P^* WP] * Patch\})$$

Below lists details about the components in the equation:

Variable	Full Name	Source	Value / Range*
C	constraints	user input (a field in the constraint dataset)	0 ~ 1
L	transition likelihood effect	calculated from user input (Attribute table - transition matrix)	0 ~ 100
W_L	weight of likelihood	calculated based on factor weight	$1-W_F$
F	factor effect	calculated from user input (factor table) and spatial location	0~100
W_F	factor weight	user input (User Interface)	0~1
P	proximity effect	calculated from user input (Attribute table - proximity) and spatial location	0~100
W_P	proximity weight	model default	0.3
PATCH	patch size constraint	calculated by comparing suitable patch size and user-defined threshold (Attribute table - patch)	0 or 1
S	suitability	calculated using variables above	0~100

** Note some of the variables are calculated from the relative numbers from user input, hence the value/range here may differ from the range required for your input.*

Change Override

While the methods described above use computational methods to change pixels and present a probable scenario land cover, there are times when the user prefers to have an exception and to override these methods to change pixels. The tool allows the user to enter an override GIS layer, which converts pixels as stated. At the simplest, supplying a land cover dataset and an override layer can be used to change specific defined polygons, in this case the scenario tool acts as a simple GIS operation tool. Currently, only one vector override layer is accepted therefore if using multiple layers they need to be combined.

Computing Transition

The final step in the procedure is to convert the pixels (land allocation). The InVEST Scenario Generator performs land cover transition by converting the suitability rasters into an array and processing each pixel converting them based on their suitability values. Starting from the cover type with the highest priority, the goal (%change) is read and pixels converted starting from the highest suitability. After each cover is processed, the converted pixels are masked so that they are not available for conversion again. Where more pixels of the same suitability are available, the tool randomly selects the available pixels from the first group (region) that it encounters.

Example As an illustration of the procedure, consider a 5 by 5 pixel landscape undergoing agricultural expansion pressure. There are a few important transitions here:

1. Forest -> Agriculture (very likely, 8)
2. Forest -> Urban (extremely likely, 10)
3. Grassland -> Agriculture (likely, 5)
4. Grassland -> Urban (very likely, 8)
5. Agriculture -> Urban (a little likely, 3)

6. Urban -> Agriculture (very seldom likely, 1)

At a stakeholder workshop (or other expert forum), the transitions above are discussed and the drivers that determine them are considered. The stakeholders likelihood values above are entered in a matrix as below.

	Forest	Ag	Urban	Grassland	Rock	Goal	Priority
Forest	0	8	10	0	0		
Ag	0	0	3	0	0	40%	10
Urban	0	1	0	0	0	100%	5
Grassland	0	5	8	0	0		
Rock	0	0	0	0	0		

After considering the driving forces (drivers) and the demand, stakeholders place a goal of 40% growth of agriculture and 100% growth of urban areas. Further, they determine that agriculture has higher priority than urban development so they are assigned values of 10 and 5 respectively. Since these values are relative, the absolute values do not matter. Where a cell (parcel) is more suitable for both agriculture and urban development, the agriculture wins due to higher priority.

This simple example considers only the likelihood matrix, but the tool has a provision for incorporating physical and environmental characteristics, which also determine the likelihood of land cover conversion. Even when both the likelihood matrix and the factors are applied, the tool allows adjustment of the contribution of each to the final suitability grid that is used for conversion.

Process explanation:

1. The objective is growth of agriculture with a goal of 40% increase (4 pixels) and growth of urban by 100% (1 pixel)
2. Cells A1, A2, B1, B2 are protected so no changes occur. The tool allows entering intermediate values of protection to control suitability for conversion
3. Agriculture and urban cells are masked because they are not converting
4. Cells with highest suitability scores (8) are converted first followed by lower suitability until goal is met or until possible cells run out
5. Proximity suitability is applied that is cell E3 is taken before C1 even though they have the same suitability
6. Cells D2 and E3 have the same suitability and proximity so one is picked at random
7. This procedure is repeated for each cover type (objective) starting from the highest priority to the lowest

Tool Process

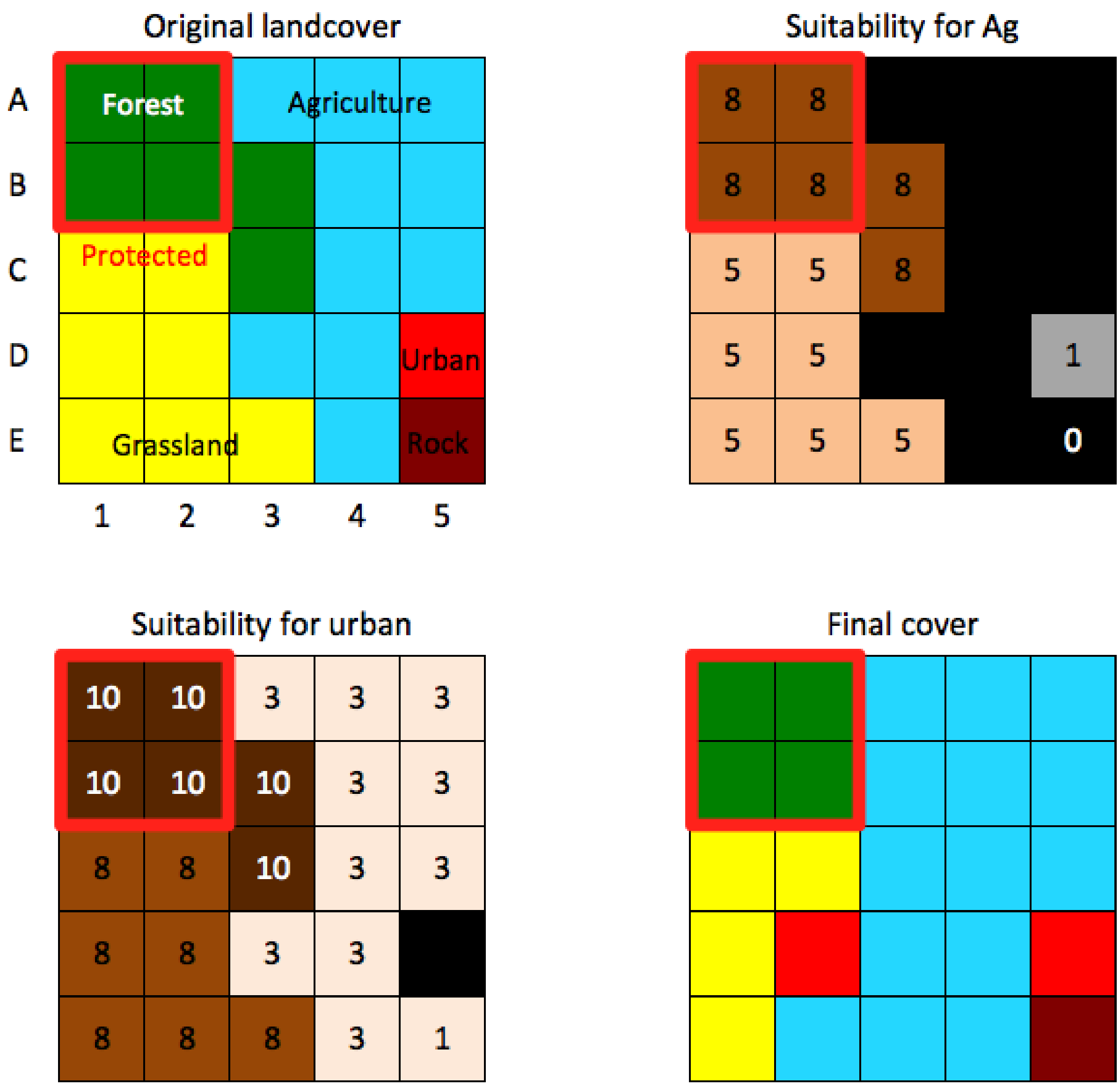
Tool Flow

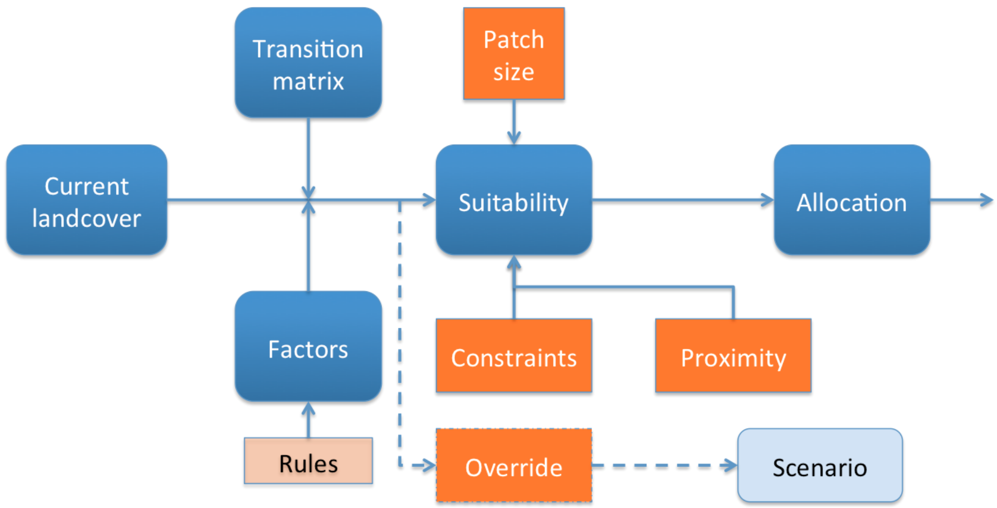
Limitations and Simplifications

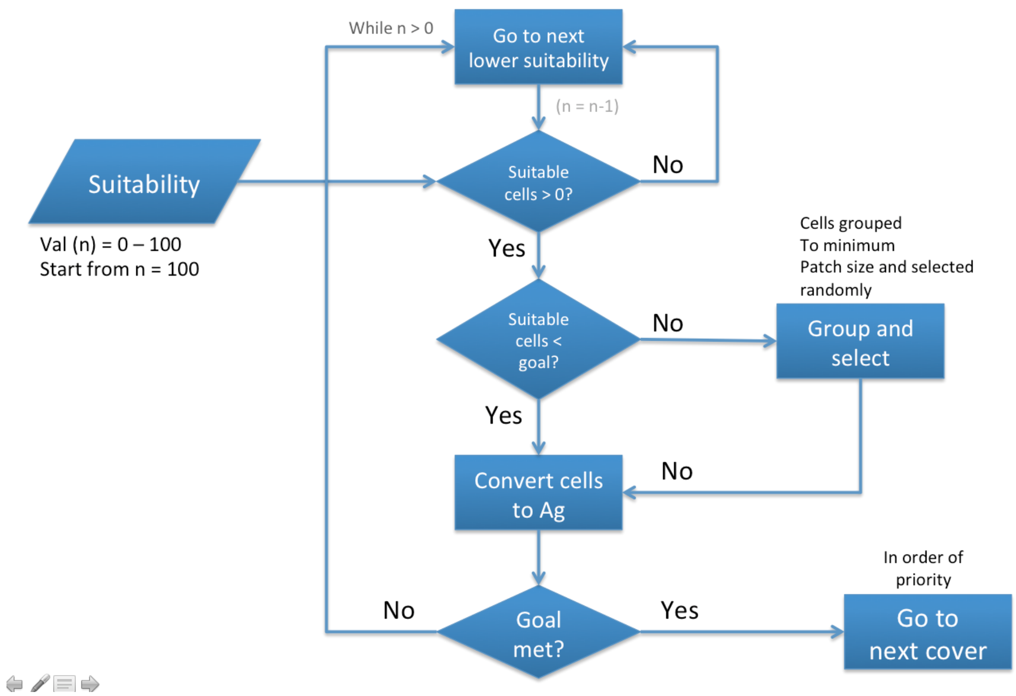
Land cover change analysis is complex and most methods only try to approximate possible futures. This model captures expert knowledge and makes an attempt at representing plausible land cover change as realistically as possible but does not predict the future land cover. One of the aims of the tool is to make it easy for stakeholders to understand how the decisions they make in building the scenarios are reflected in the map produced.

Following are some limitations/assumptions:

1. This model assumes that a cover type is either growing or shrinking but not both. In reality, conversion takes place in both directions but for simplicity, only one direction is assumed.


Original landcover
A
B
C
D
E
Forest
Agriculture
Protected
Urban
Grassland
Rock
1
2
3
4
5
Suitability for Ag
Suitability for urban
Final cover


Transition
matrix
Patch
size
Current
landcover
Suitability
Allocation
Factors
Constraints
Proximity
Rules
Override
Scenario


While n > 0
Go to next
lower suitability
(n = n-1)
Suitability
Val (n) = 0 – 100
Start from n = 100
Suitable
cells > 0?
No
Yes
Cells grouped
To minimum
Patch size and selected
randomly
Suitable
cells <
goal?
No
Group and
select
Yes
Convert cells
to Ag
No
In order of
priority
No
Goal
met?
Yes
Go to
next cover

2. This tool assumes a single step transition from the beginning landcover to the scenario landcover. In reality, these changes could be stepwise with different patterns at each step.
3. Stakeholder values are likely to be more reliable for near future scenarios but not for longer term ones. Therefore, it is advisable to stay with near future.
4. Currently, this tool only processes covers that are growing and disregards the shrinking covers. For example, even though a percentage change may be entered as -5%, the pressure of the growing covers always override. In a future revision this will be addressed.
5. The sequence of land transition is solely based on priority, which means even if a land parcel is highly suitable for agriculture but barely suitable for urban, it may still be assigned to future urban land as long as urban has a higher priority.

Current Issues, and Ways to Avoid Errors

We are constantly improving and developing new functions for Scenario Generator. However, below are some issues reported by users, which we will correct in future releases. For now, we provide some possible tricks to avoid them before the updated new version is ready.

Issue – Error if you run the model without "Specify Transitions" when the scenario has new LULC types.
: **Error Message:**

  Traceback (most recent call last):

  File "invest_natcapiuiexecutor.pyc", line 555, in runModel

  File "invest_natcapscenario_generatorscenario_generator.pyc", line 1148, in execute

  KeyError: 2 (2 is the LULC ID that does not exist in the baseline LULC map)

  None

  Suggestion: Check the "Specify Transition" option and run the model with a transition matrix when new land covers are expected.

Issue – Error If the user checked proximity but put nothing into the proximity column: Error Message:

  Traceback (most recent call last):

  File "invest_natcapiuiexecutor.pyc", line 555, in runModel

  File "invest_natcapscenario_generatorscenario_generator.pyc", line 946, in execute

  ValueError: invalid literal for int() with base 10: ''

  Suggestion: If proximity is not applied to a certain layer, put in 0 instead of leave it blank. Also, if you are not using the proximity element, don't check the box.

Issue – Performance issues with large landscapes. Users have reported memory errors.
: **Suggestion:** Downsample your map and re-run with coarse resolution, or breakdown your study area and build separate scenarios for them.

Issue – The randomization algorithm seems to have a problem. It has been noted in some cases that pixel conversion in no
: **Suggestion:** Sorry for that... Before we correct the code, probably you can set a lower goal to avoid the random allocation process in the tool. (And maybe use random points tool in GIS software to generate your extra new land cover)

Issue – When introducing a new landcover type, the raster attribute table is not built automatically.
: **Suggestion:** Build a raster attribute table for your scenario map, then the raster can symbolize correctly.

Issue – The override function should be able to work independently of everything else. Currently, if you do not do transiti

Suggestion: For now, if the override function is the only process you need to generate scenarios, try using GIS tools instead of using Scenario Generator

Issue – Error when table names include non-ASCII characters Error Message:

UnicodeEncodeError: 'ascii' codec can't encode character u'xeb' in position 726: ordinal not in range(128)

Suggestion: Use the 26 English letters only for your variable/land cover names.

7.3.4 Data Needs

1. **Result suffix (optional):** The value entered here will be used as a suffix for your results. Only one character is allowed. Adding a unique suffix will avoid overwriting previous results files.
2. **Base Land cover:** Land cover data in raster format. While the number of land cover classes can be unlimited, for this analysis it gets confusing for stakeholders and becomes problematic to process a large number of land cover classes. Its preferable to keep them under 20. Stakeholders should be able to describe each of the covers especially those that are transitioning.
3. **Landcover transition table:** A csv table that contains the transition likelihoods on a scale of 0 to 10 where 0 indicates no likelihood of change and 10 indicates full likelihood of change. The rows indicate the land cover types. For each land cover type in the row, there is a matching field named <cover id> where the cover id matches the id in the row as shown in the example below. Additional fields include:
    1. Priority: If the user has priority for the cover types, they should be entered here. Otherwise, the Compute Priority option should be used to populate this field. The tool will allocate pixels for the cover types with a higher priority before those with lower priority.
    2. Percent Change: This shows the quantity of change and should be a positive integer. We currently only model cover growth, negative values will be ignored. This is a limitation.
    3. Proximity: If proximity suitability is to be applied to this cover type, enter the proximity distance; else leave it as 0. The proximity distance is a value in meters that indicates how far the effect of self proximity goes. For example, fields that are within 10km of small scale agriculture may be likely to be converted to agriculture if they are suitable, but after this distance the effect of proximity disappears. In this case, a value of 10000 should be entered.
    4. Area Change (optional): For cover types that do not already exist, percentage change cannot be used. To introduce a new cover, enter the new quantity in hectares.
    5. Patch ha (optional): This is an optional value that indicates the minimum size of a patch that is suitable for the cover to be allocated the parcel. If not entered, a default value of 1 pixel is used.

Id	Name	Short	1	2	3	4	Percent Change	Area Change	Priority	Proximity	Patch (ha)
1	Grassland	Grassland	0	4	0	1	0	0	0	0	0
2	Agriculture	Agriculture	0	0	0	0	0	8000	8	5000	0
3	Forest	Forest	0	8	0	1	0	0	0	0	0
4	Bareland	Bareland	0	0	0	0	25%	0	5	10000	

In the table above, there is growth in agriculture and bare land at the expense of grassland and tropical forest. The likelihood of tropical forest transitioning to agriculture is rated 8 while grassland to agriculture is rated 4 therefore when converting pixels to agriculture, the forest pixels are converted before grassland pixels (see assumptions). Similarly, when converting pixels, the goal of agriculture is satisfied before bare land because it has higher priority.

1. **Priority matrix (optional):** To rank the cover types for conversion, the priority of land covers are calculated using the multi-criteria evaluation approach, applying pairwise comparison with the analytic hierarchy process.

   If the user does not want to use this approach they can manually enter the priority into the respective table. The matrix should follow the format below, where matrix values are entered from column 3. The first two columns are reserved for descriptive values (Record number and the item) and the last column is reserved for the PRIORITY. The tool computes the priority and populates this column. Only the lower half of the diagonal should be filled and the diagonal cells should contain 1s. The names of the items are not crucial, but they must be in the same order as they are in the rows.

id	Name	Cover A	Cover B	Cover C	Priority
1	Cover A	1			
2	Cover B	0.5	1		
3	Cover C	0.1	5	1	

2. **Land suitability factors (optional):** This table lists the factors that determine suitability of the land cover for change. Each factor lists a layer, which defines the suitability. Given that the same factor can have different implications for different objectives, users can enter more than one layer for each cover (objective). If this table is not provided, these factors will not be used and only the transition likelihood table above will be used. It is strongly advised to include factors. The following are the required fields:

   1. Cover ID: The ID of the land cover affected by the factor
   2. Short Name: Short name of the land cover type
   3. Factorname: The name of the factor. This should be a single short name for identifying the factor and unique for the factor. No spaces allowed
   4. Layer: The name of the GIS feature class with the features of the factor. For example roads.shp. Area (as opposed to lines and points) datasets are currently only given as vectors (e.g. shapefiles). If given as vector then 'suitfield' (with values in the range 0-100) must be specified in the file's attribute table.
   5. Wt: This is the weight of the factor. When factors are combined, this weighting is applied.
   6. Suitfield: This identifies the field in the polygon layer that contains the suitability value. The field values should be integers number between 0 and 100 with 0 being unsuitable and 100 being very suitable.
   7. Dist: The distance of influence of the factor e.g. the distance from the roads. This tool uses just one distance for all the features. The polygon features do not use this field. Distance should be in the units of the landcover dataset (assumed meters).

id	Factorname	Layer	Wt	Suitfield	Dist	Cover
2	roads	Roads.shp	5		10000	Smallscl
3	majorroads	majorroads.shp	2		6000	Largescl
4	elevation	Elevation.shp	3	suit		Agric
5	roads	Roads.shp	5			Largescl
6	elevation	Elevation.shp	8	suit		Largescl

1. **Specify transitions:** This is an option that determines whether the likelihood probability matrix should be used. If this is not checked the probability matrix is ignored. This value is entered on the interface.
2. **Use factors:** This is an option that determines whether the factors should be used. If this is not checked the factors are ignored. This value is entered on the interface.
3. **Factor weight:** The factor weight is a value between 0 and 1 which determines the weight given to the factors vs. the expert opinion likelihood rasters. For example, if a weight of 0.3 is entered then 30% of the final

suitability is contributed by the factors and the likelihood matrix contributes 70%. This value is entered on the tool interface.

4. **Constraints Layer (optional):** This is a vector layer which indicates the parts of the landscape that are protected of having constraints to land cover change. The layer should have one field named 'protlevel' with a value between 0 and 1 where 0 means its fully protected and 1 means its fully open to change.
5. **Change override layer (optional):** This is a vector (polygon) layer with land cover types in the same scale and projection as the input land cover. This layer is used to override all the changes and is applied after the rule conversion is complete.

7.3.5 Interpreting Results

Final Results

Final results are generated in the "output" folder of the workspace for this module. Typically the tool is run several times changing the values until and acceptable scenario map is produced.

1. scenario.tif - This is the new landcover data created. Load this data and compare with the original landcover.
2. scenario-output-summary.html - a html file which shows the land cover transitions and a graphical representation of the changes.

Intermediate results

The intermediate folder contains the intermediate files used in the model run.

7.3.6 References

1. Carver, S. J. (1991) Integrating multi-criteria evaluation with geographical information systems International Journal of Information Systems 5 (3) 321-339
2. Collins, M.G., Steiner, F.R. and Rushman, M. J. (2001) Environmental Management 28 (5) 611-621
3. Malczewski, J. (2004) GIS-based land-use suitability analysis: a critical overview Progress in Planning 62 3-65
4. Saaty, T.L (1977) A Scaling Method for Priorities in Hierarchical Structures Journal of Mathematical Psychology 15, 234-281
5. Saaty, T. L. (2008) Decision Making with the analytic hierarchy process International Journal of Services Sciences 1(1) 83-98
6. Veldkamp, A. and Lambin, E.F. (2001) Predicting Land-Use change Agriculture Ecosystems and Environment.

7.3.7 Appendix

The Saaty 9 Point Continous Scale

1	Equal Importance	Two activities contribute equally to the objective
2	Weak or slight	
3	Moderate importance	Experience and judgment slightly favor one activity over another
4	Moderate plus	
5	Strong importance	Experience and judgment strongly favor one activity over another
6	Strong plus	
7	Very strong or demonstrated importance	An activity is favored very strongly over another; its dominance demonstrated in practice
8	Very very strong	
9	Extreme importance	The evidence favoring one activity over another is of the highest possible order of affirmation

7.4 Scenario Generator: Proximity Based

7.4.1 Summary

The proximity-based scenario generator creates a set of contrasting land use change maps that convert habitat in different spatial patterns. The user determines which habitat can be converted and what they are converted to, as well as type of pattern, based on proximity to the edge of a focal habitat. In this manner, an array of land-use change patterns can be generated, including pasture encroaching into forest from the forest edge, agriculture expanding from currently cropped areas, forest fragmentation, and many others. The resulting land-use maps can then be used as inputs to InVEST models, or other models for biodiversity or ecosystem services that are responsive to land use change.

7.4.2 Introduction

In order to understand the change in biodiversity and ecosystem services (BES) resulting from change in land-use, it is often helpful to start with a scenario or a set of scenarios that exhibit different types of land-use change. Because many of the relationships between landscapes and BES are spatially-explicit, a different pattern of habitat conversion for the same total area of habitat converted can lead to very different impacts on BES. This proximity-based scenario generator creates different patterns of conversion according to user inputs designating focal habitat and converted habitat, in contrast to but potentially complementing the InVEST rule-based scenario generator that creates maps of land-use change according to user-assigned probabilities that certain transitions will occur. Thus, the intent of the InVEST proximity-based scenario generator is not to forecast actual predicted patterns of expansion, but rather to develop different patterns of land use change in order to examine the relationship between land-use change and BES, and how the relationship may differ depending on land use change assumptions.

7.4.3 The model

The tool can generate two scenarios at once (nearest to the edge and farthest from the edge of a focal habitat), for a conversion to particular habitat type for a given area. To convert to different habitat types, different habitat amounts, or to designate different focal habitats or converted habitats, the tool can be run multiple times in sequence.

How it works

Three types of landcover must be defined: 1) *Focal landcover* is the landcover(s) that set the proximity rules from which the scenarios will be determined. The scenario generator will convert habitat from the edge or toward the edge of patches of these types of landcover. This does not mean it will convert these land-covers, only that it will measure distance to or from the edges in designating where the conversion will happen. 2) *Convertible landcover* is the landcover(s) that can be converted. These could be the same as the focal landcover(s), a subset, or completely different. 3) *Replacement landcover* is the landcover type to which the convertible landcovers will be converted. This can only be one landcover type per model run.

Two scenarios can then be run at a time: 1) *Nearest to edge* means that convertible landcover types closest to the edges of focal landcovers will be converted to the replacement landcover. 2) *Farthest from edge* means that convertible landcover types furthest from the edges of focal landcover types will be converted to the replacement landcover. If this scenario is chosen, the user can designate in how many steps the conversion should occur. This is relevant if the focal landcover is the same as the convertible land cover because the conversion of the focal landcover will create new edges and hence will affect the distance calculated from the edge of that landcover. If desired, the conversion can occur in several steps, each time converting the farthest from the edge of the focal landcover, causing a fragmentary pattern.

Below are some examples of the types of scenarios that can be generated by manipulating these basic inputs, using the land-cover in the sample data that ship with this model. This landcover is from MODIS, using the UMD classification (Friedl et al. 2011), which follows the following scheme: 1 – Evergreen needleleaf forest; 2 – Evergreen broadleaf forest; 3 – Deciduous needleleaf forest; 4 – Deciduous broadleaf forest; 5 – Mixed forest; 6 – Closed shrublands; 7 – Open shrublands; 8 – Woody savannas; 9 – Savannas; 10 – Grasslands; 12 – Croplands; 13 – Urban and built-up; 16 – Barren or sparsely vegetated.

Expand agriculture from forest edge inwards:

focal landcover codes: 1 2 3 4 5

convertible landcover codes: 1 2 3 4 5

replacement landcover code: 12

check “Convert From Edge”

number of steps toward conversion: 1

Expand agriculture from forest core outwards:

focal landcover codes: 1 2 3 4 5

convertible landcover codes: 1 2 3 4 5

replacement landcover code: 12

check “Convert Toward Edge”

number of steps toward conversion: 1

Expand agriculture by fragmenting forest:

focal landcover codes: 1 2 3 4 5

convertible landcover codes: 1 2 3 4 5

replacement landcover code: 12

check “Convert Toward Edge”

number of steps toward conversion: 10 (or as many steps as desired; the more steps, the more finely fragmented it will be and the longer the simulation will take)

Expand pasture into forest nearest to existing agriculture:

focal landcover codes: 12

convertible landcover codes: 1 2 3 4 5

replacement landcover code: 10

check "Convert From Edge"

number of steps toward conversion: 1

7.4.4 Data needs

The only required input data to run the proximity-based scenario generator is a base land-use/land-cover map and user-defined land cover codes pertaining to this base map to designate how the scenarios should be computed.

1. Base land-use/cover map (required). This is the map that will be modified in the generation of the desired scenarios. All pixels in this map (that overlap with the area of interest, if included) other than the pixels that are converted will remain the same.

    Name: file can be named anything (scenario_proximity_lulc.tif in the sample data)

    Format: standard GIS raster file (e.g., ESRI GRID or IMG), with a column labeled 'value' that designates the LULC class code for each cell (integers only; e.g., 1 for forest, 10 for grassland, etc.)

1. AOI – Area of Interest (optional). If change is only desired in a subregion of the broader land-use/land-cover map, the user may designate this area of interest. Prior to scenario generation, the map will be clipped to the extent of this vector.

   Name: file can be named anything (scenario_proximity_aoi.shp in the sample data)

   Format: vector (polygon) file

2. Max area to convert (ha): enter the maximum numbers of hectares to be converted to agriculture. This is the maximum because due to the discretization of area of pixels, the number of pixels closest to but not exceeding this number will be converted.

3. Focal Landcover Codes: enter the LULC code(s) for the land cover types from which distance from edge should be calculated. If multiple values, they should be separated by spaces.

4. Convertible Landcover Codes: enter the LULC code(s) for the land cover types that are allowed to be converted to agriculture in the simulation. If multiple values, they should be separated by spaces.

5. Replacement Landcover Code: enter an integer that corresponds to the LULC code to which habitat will be converted. If there are multiple LULC types that are of interest for conversion, this tool should be run in sequence, choosing one type of conversion each time. A new code may be introduced if it is a novel land-use for the region or if it is desirable to track the expanded land-use as separate from historic land-use.

6. Check boxes: types of scenarios to generate

   (a) Convert farthest from edge: land cover type(s) designated as "convertible" that are farthest from the edge of any land cover type designated as "focal" will be converted. Convertible land covers and habitat of interest land-covers may be the same, or a subset of one another, or they can be different. If they are the same, the number of steps for conversion should be specified, because the conversion of habitat within the focal land cover will create new habitat edge, resulting in a completely different pattern of conversion depending on how many steps are chosen.

   (b) Convert nearest to edge: land cover type(s) designated as "convertible" that are nearest to the edge of any land cover type designated as "focal" will be converted. As for the previous scenario, convertible land covers and habitat of interest land-covers may be the same, or a subset of one another, or they can be different.

7. Number of Steps in Conversion: enter an integer for the number of steps the simulation should take to fragment the habitat of interest in the fragmentation scenario. Entering a 1 means that all of the habitat conversion will occur in the center of the patch of the habitat of interest. Entering 10 will be fragmented according to a pattern

of sequentially converting the pixels furthest from the edge of that habitat, over the number of steps specified by the user.

7.4.5 Running the model

The model is available as a standalone application accessible from the install directory of InVEST (under the subdirectory invest-3_x86/invest_scenario_gen_proximity.bat).

Viewing Output from the Model

Upon successful completion of the model, a file explorer window will open to the output workspace specified in the model run. This directory contains an output folder holding files generated by this model. Those files can be viewed in any GIS tool such as ArcGIS, or QGIS. These files are described below in Section Interpreting Results.

7.4.6 Interpreting Results

Final Results

Final results are found in the *Workspace* folder within the specified for this module.

- **natcap.invest.ag_expansion-log**: Each time the model is run, a text (.txt) file will appear in the *Output* folder. The file will list the parameter values for that run and will be named according to the service, the date and time, and the suffix.
- **nearest_to_edge _<suffix>.tif:** LULC raster for the scenario of conversion nearest to the edge of the focal habitat
- **farthest_from_edge_<suffix>.tif:** LULC raster for the scenario of conversion farthest from the edge of the focal habitat
- **nearest_to__edge_ stats_<suffix>.csv:** table listing the area (in hectares) and number of pixels for different land cover types converted for the scenario of conversion nearest to the edge of the focal habitat
- **farthest_from_edge_ stats_<suffix>.csv:** table listing the area (in hectares) and number of pixels for different land cover types converted for the scenario of conversion nearest to the edge of the focal habitat

Intermediate Results

You may also want to examine the intermediate results. These files can help determine the reasons for the patterns in the final results. They are found in the *intermediate_outputs* folder within the *Workspace* specified for this module.

- **{farthest_from_/nearest_to}_edge_distance_<suffix>.tif: map** of This raster shows the distance (in number of pixels) of each pixel to the nearest edge of the focal landcover

7.4.7 Sample Script

The following script is provided to demonstrate how the scenarios described in Section "How It Works" can be composed into a single script that's callable from the InVEST Python API.

import natcap.invest.scenario_generator_proximity_based

edge_args = {

u'aoi_uri': u'C:/Users/Rich/Documents/svn_repos/invest-sample-data/scenario_proximity/scenario_proximity_aoi.shp', u'area_to_convert': u'20000.0',

u'base_lulc_uri': u'C:/Users/Rich/Documents/svn_repos/invest-sample-data/scenario_proximity/scenario_proximity_lulc.tif',

u'convert_farthest_from_edge': False,

u'convert_nearest_to_edge': True,

u'convertible_landcover_codes': u'1 2 3 4 5',

u'focal_landcover_codes': u'1 2 3 4 5',

u'n_fragmentation_steps': u'1',

u'replacment_lucode': u'12',

u'results_suffix': 'edge',

u'workspace_dir': u'C:\\\\Users\\\\Rich/Documents/scenario_proximity_workspace',

}

core_args = {

u'aoi_uri': u'C:/Users/Rich/Documents/svn_repos/invest-sample-data/scenario_proximity/scenario_proximity_aoi.shp',

u'area_to_convert': u'20000.0',

u'base_lulc_uri': u'C:/Users/Rich/Documents/svn_repos/invest-sample-data/scenario_proximity/scenario_proximity_lulc.tif',

u'convert_farthest_from_edge': True,

u'convert_nearest_to_edge': False,

u'convertible_landcover_codes': u'1 2 3 4 5',

u'focal_landcover_codes': u'1 2 3 4 5',

u'n_fragmentation_steps': u'1',

u'replacment_lucode': u'12',

u'results_suffix': 'core',

u'workspace_dir': u'C:\\\\Users\\\\Rich/Documents/scenario_proximity_workspace',

}

frag_args = {

u'aoi_uri': u'C:/Users/Rich/Documents/svn_repos/invest-sample-data/scenario_proximity/scenario_proximity_aoi.shp',

u'area_to_convert': u'20000.0',

u'base_lulc_uri': u'C:/Users/Rich/Documents/svn_repos/invest-sample-data/scenario_proximity/scenario_proximity_lulc.tif',

u'convert_farthest_from_edge': True,

u'convert_nearest_to_edge': False,

u'convertible_landcover_codes': u'1 2 3 4 5',

u'focal_landcover_codes': u'1 2 3 4 5',

u'n_fragmentation_steps': u'10',

u'replacment_lucode': u'12',

u'results_suffix': 'frag',

u'workspace_dir': u'C:\\\\Users\\\\Rich/Documents/scenario_proximity_workspace',

}

ag_args = {

u'aoi_uri': u'C:/Users/Rich/Documents/svn_repos/invest-sample-data/scenario_proximity/scenario_proximity_aoi.shp',

u'area_to_convert': u'20000.0',

u'base_lulc_uri': u'C:/Users/Rich/Documents/svn_repos/invest-sample-data/scenario_proximity/scenario_proximity_lulc.tif',

u'convert_farthest_from_edge': False,

u'convert_nearest_to_edge': True,

u'convertible_landcover_codes': u'12',

u'focal_landcover_codes': u'1 2 3 4 5',

u'n_fragmentation_steps': u'1',

u'replacment_lucode': u'12',

u'results_suffix': 'ag',

u'workspace_dir': u'C:\\Users\\Rich/Documents/scenario_proximity_workspace',

}

if __name__ == '__main__':

natcap.invest.scenario_generator_proximity_based.execute(edge_args)

natcap.invest.scenario_generator_proximity_based.execute(core_args)

natcap.invest.scenario_generator_proximity_based.execute(frag_args)

natcap.invest.scenario_generator_proximity_based.execute(ag_args)

7.5 Scenario Hub

7.5.1 Introduction

The Scenario Hub (scenariohub.net) is a web-based portal that aims to:

1. Provide standard questionnaire to collect information and generate growth estimates (under development) from experts and stakeholders.
2. Help users document scenarios and prepare input data for Scenario Generator.
3. Encourage information sharing on different projects and scenarios, and provide a case pool for Scenario Generator users.

We suggest users refer to the Scenario Generator user's guide for the modeling part and the use of each estimate.

7.5.2 How it works

Registration

Scenario Hub requires a user account to access its full function. Use the Auth Code / registration code **NATCAP2015** to create a new account. After you verified your email address, you can sign in with your account and create scenarios.

View Projects

By click on project names, you can view other people's projects. Click on the scenario name to expand the associated scenario profile.

Create A Project

Enter the name, time horizon, and general description to create a new project. You can also choose whether your project is visible to other users or not. Once a project created, you won't be able to edit the project information.

Most of the inputs in this section are not required, nor will they be used to build up your Scenario Generator input files. However, we do suggest that you post relevant information here, so that you can keep a better record of your project, and other users in the community will benefit from your case as well.

Create Scenarios

You can create and edit your scenarios in your own project page. Components required to build you scenario profile include:

A. General Information:

1. **Scenarios:** The names of scenarios you will develop for the project. You need to enter a scenario name here to expand the estimates sections (discussed later).
2. **Scenario Description (optional):** We suggest you put the detailed storyline here, which will help others (and yourself) understand the assumptions and reasons behind your estimates.

B. Estimates:

Expand your estimates section by clicking the scenario name you wish to edit. The web interface will use the information you post in this section to build the csv tables used by Scenario Generator.

1. **Landcovers:** The list of landcover types that you are going to model. Including both LULCs in current status and new LULCs in your scenario. No duplicated entries. Required attributes for each entry include LULC ID (as in the raster dataset), Landcover Name, and whether the type exist in current/baseline map or not. If the cover type exists only in your scenario, check the box for "New Cover".
2. **Drivers:** The activities/processes that contribute to or lead to future land cover change. Drivers won't be used for the modeling purpose, but they will help you clarify the underlying process and possible linkage between current activities and future land changes. Therefore, we require user input of drivers when add a land transition to force the consideration of the process and linkage.
3. **Transitions:** Possible land transitions under the scenario. The information you entered here will be used to build the transition matrix. Required attributes for entries include:

1. *From Cover:* the LULC type going to decrease in the transition. The site will provide you a drop-down list with all the landcover types you entered in step 1), so that you can select from them. However, once you put an LULC to the "From Cover" field, it will disappear from the "To Cover" drop-down list, and vice versa. That is because the Scenario Generator currently only support growth change modeling, and does not allow an LULC experiencing both growth and decrease.
2. *To Cover:* the LULC type going to increase in the transition.
3. *Driver:* the driving force of the transition. You will choose from the list of drivers entered in the previous section.

4. *Likelihood:* the likelihood for such transition to happen.

4. **Cover Growth:** Inputs will be used to build the land transition attribute table

1. *Cover:* The cover that will grow in the scenario. The web application will automatically add "To Cover" here after you add a transition.

2. *Increase:* The amount of future growth of the LULC in either percentage or hectare. It will be used as the goal for land growth allocation model. (See Scenario Generator for more details.)

3. *Units:* the unit of the increase mentioned above. Note that if the cover is a New Cover, you need to select "ha" as the unit. Otherwise please report the increase in percentage.

4. *Priority:* The relative value here will decide which landcover will have its transition modeled earlier. A larger number indicates higher priority, and the cover will be modeled before those with lower priority. Since the value is relative here, the exact number does not matter.

5. *Proximity (m) (optional):* The distance limit of proximity effect. Leave it blank or put in 0 if no proximity effect is expected.

6. *Patch (ha) (optional):* The minimum patch size required for the transition. Any suitable patches smaller than the size will not be converted.

5. **Factors (optional):** Environmental/physical variables that affect the physical suitability for a cover type. The information entered will be used to build the factor table for Scenario Generator.

1. *Factor Name.*

2. *Landcover:* The target landcover whose growth will be affected by the factor.

3. *GIS Layer:* GIS layer file name of the factor.

4. *Distance of influence (m):* For line and point factor features, enter the distance of influence in meters.

5. *Suitability Field:* For polygon factors, enter the field indicating the suitability level. The range of the value is 0~100.

6. *Weight:* The relative weight of this factor comparing to other factors affecting the same target layer.

7. *Data Type:* Select from point, line, or polygon.

6. **Constraints:** Areas that land changes area restricted / discouraged. The information entered here serves as a reference. When running Scenario Generator, you will need to provide the Constraints data through the user interface.

After filling in all the sections, you can download the attributes table and factors table by clicking the buttons at the bottom of the page. Those two tables are in csv format and suits the input requirement for the Scenario Generator.

7.5.3 Tips

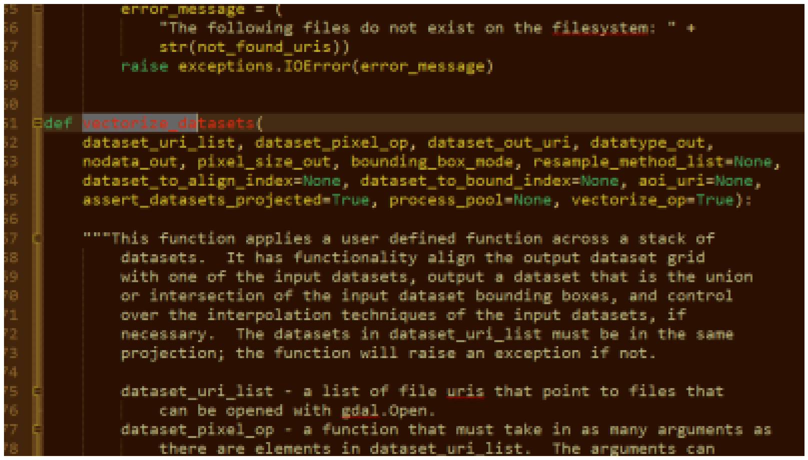
1. If your input does not show up after you click "save" : try to refresh the page.

2. The likelihood of transition only allows you to choose from limited options. If you want more variation, just download the attribute table and edit the csv file.

3. For the Growth Priority, you can either enter values from 1 ~ 100, or 0.01 ~ 1. However, do not put in a priority power of 80 among a group of 0.xx, otherwise your 0.xx values might be adjusted to zeros hence showing no difference. Also, remember that higher value indicates higher priority here.

7.6 InVEST Scripting Guide and API

7.6.1 Summary



7.6.2 Introduction

While there is a user interface for all the InVEST models, we also provide a Python application programming interface (API) for it. This includes the main entry points for the InVEST ecosystem service models, and our underlying high performance geoprocessing and routing pipeline. At the moment the documentation is limited, but we do support the installation of the API standalone through our Bitbucket Wiki site: https://bitbucket.org/natcap/invest-natcap.invest-3/wiki/Scripting%20InVEST%20on%20Windows

This documentation will be updated in more detail in the future. Programming support is also available on our user forums: http://forums.naturalcapitalproject.org/.

Part III

Acknowledgements

CHAPTER

EIGHT

ACKNOWLEDGEMENTS

8.1 Data sources

The Marine InVEST development team would like to acknowledge the following sources for data that are provided with the models:

WAVEWATCH III model hindcast reanalysis results are from NOAA's National Weather Service

ETOPO1 was developed by and is available from NOAA's National Geophysical Data Center (NGDC). (Amante, C. and B. W. Eakins, ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24, 19 pp, March 2009).

The Global Self-consistent, Hierarchical, High-resolution Shoreline Database (GSHHS) is developed and maintained by Paul Wessel, SOEST, University of Hawai'i, Honolulu, and Walter H.F. Smith, NOAA Geosciences Lab, National Ocean Service, Silver Spring, MD. It can be accessed via NOAA's National Geophysical Data Center (NGDC). (Wessel, P., and W. H. F. Smith, A Global Self-consistent, Hierarchical, High-resolution Shoreline Database, J. Geophys. Res., 101)

British Columbia Shorezone Data are provided courtesy of the Province of British Columbia, Ministry of Natural Resource Operations, GeoBC Division. The data used for this model is a snapshot in time. For the most current coastal resource and shorezone data please visit http://www.geobc.gov.bc.ca.

Gridded Population of the World Version 3 (GPWv3) data are provided courtesy of the Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro Internacional de Agricultura Tropical (CIAT). 2005. Gridded Population of the World Version 3 (GPWv3): Population Grids. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), CIESIN, Columbia University. Available at http://sedac.ciesin.columbia.edu/gpw.

Department of Fisheries and Oceans Canada provides commercial fishery data layers (groundfish trawl and longline, salmon troll, and shrimp trawl) collected from 1993-95 through interviews with fisheries officers for the West Coast of Vancouver Island, British Columbia. Data are publicly available through GeoBC, the geographic information clearinghouse for British Columbia Canada.

Habitat layers are available through British Columbia Marine Conservation Analysis (BCMCA) and GeoBC.

8.2 Individuals and organizations

We would also like to thank the following individuals and organizations for support, guidance, great collaborations, and fun! This is not an exhaustive list.

The Gordon and Betty Moore Foundation

The West Coast Aquatic Management Board

NOAA Fisheries, Northwest Fisheries Science Center

Gretchen Daily, Peter Kareiva, Taylor Ricketts, Steve Polasky, and Jon Foley

Our Technical Working Group: Mary Ruckelshaus, Anne Guerry, Katie Arkema, Greg Guannel, CK Kim, Mike Papenfus, Jodie Toft, Gregg Verutes, Joey Bernhardt, Apollo Qi, Jeremy Davies, Heather Tallis, Steve Polasky, Mark Plummer, Phil Levin, Bill Labiosa, Francis Chan, Guy Gelfenbaum, Peter Ruggiero, Andre Punt, Melanie McField, Ben Halpern, Sarah Lester, Malin Pinsky, Mike Beck, Barry Gold, Kai Chan

Others: Matt Marsik, Spencer Wood, Dave Sutherland, Andrew Day, Laura Loucks, Trudy Warner, Kevin Head, Roger Bedard, Jim Regetz, Dan Holland, Jameal Samhouri .. primerend

Part IV

PDF Version of the User's Guide

Download a pdf version of the InVEST documentation.