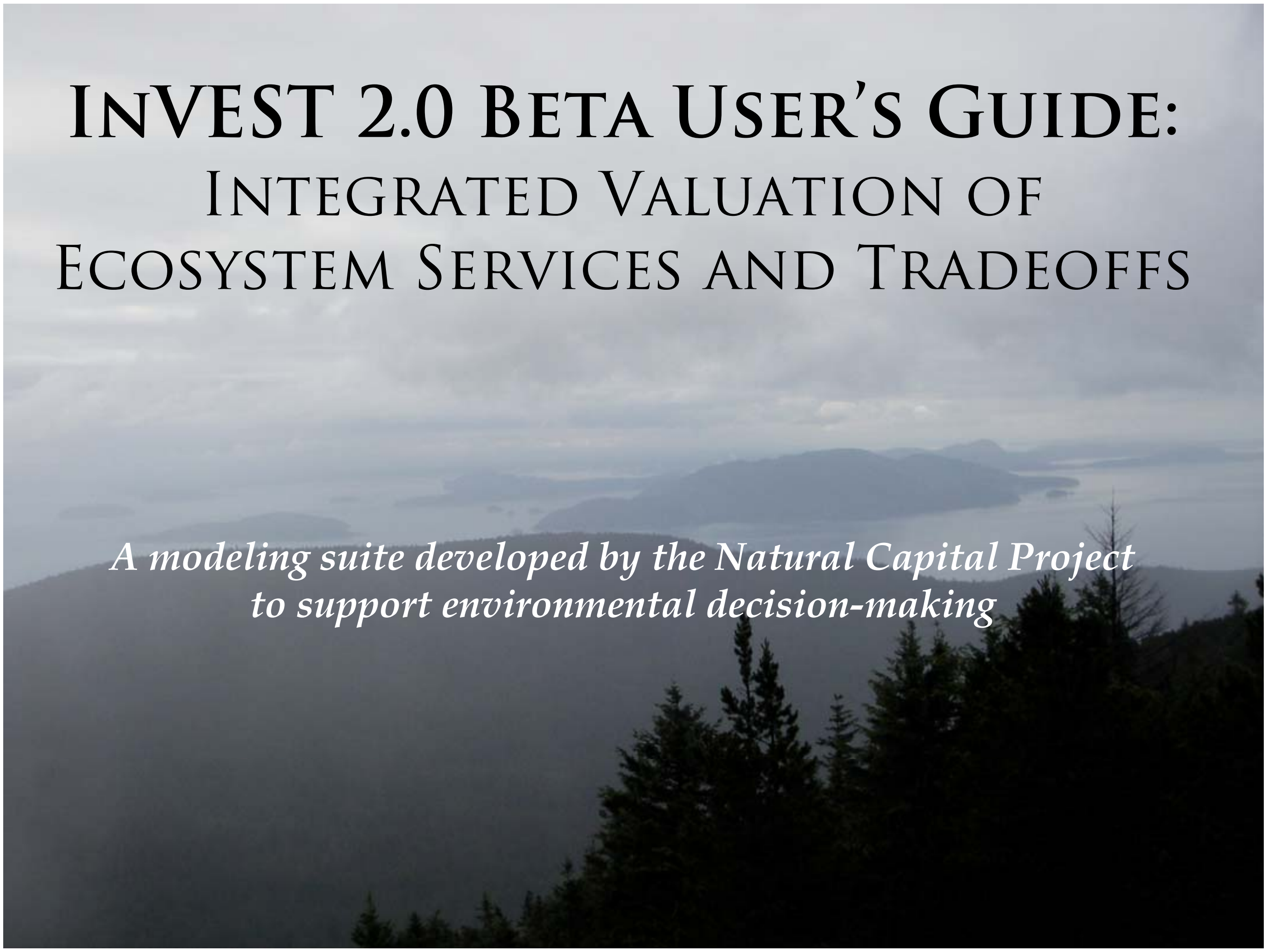
InVEST 2.0 Beta User's Guide:
Integrated Valuation of Ecosystem Services and Tradeoffs



A modeling suite developed by the Natural Capital Project to support environmental decision-making

Editors: Heather Tallis, Taylor Ricketts, and Anne Guerry

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, Gregg Verutes, Choong-Ki Kim, Gregory Guannel, Michael Papenfus, Jodie Toft, Matthew Marsik, and Joey Bernhardt.

Citation: Tallis, H.T., Ricketts, T., Guerry, A.D., Nelson, E, Ennaanay, D., Wolny, S, Olwero, N., Vigerstol, K., Pennington, D., Mendoza, G.,.Aukema, J., Foster, J., Forrest, J., Cameron, D, Lonsdorf, E., Kennedy, C., Verutes, G., Kim, C.K., Guannel, G., Papenfus, M., Toft, J., Marsik, M., and Bernhardt, J. 2011. InVEST 2.0 beta User's Guide. The Natural Capital Project, Stanford.

TABLE OF CONTENTS

THE NEED FOR A NEW TOOL

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), and 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 better align ecosystem conservation with economic forces, 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 development of the models and will release new, updated versions as they become available. The next release is planned for May 2011.

Who should use InVEST?

InVEST is designed to inform decisions about natural resource management. 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 evaluating these trade-offs. For example, government agencies could use InVEST to help determine how to manage lands, coasts, and marine areas to provide an optimal mix 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 timber 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 preserved.

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?

Introduction to InVEST

The InVEST 2.0 Beta package described in this guide includes the first models for quantifying, mapping, and valuing the benefits provided by marine systems. Four marine models are included in this release (wave energy, coastal exposure, marine fish aquaculture, and aesthetic views). Additional marine models (including those for fisheries, recreation and tourism, protection from erosion/flooding, and carbon storage and sequestration) will be included in the next release in May 2011.

Terrestrial and freshwater models within InVEST 2.0 Beta include those for carbon sequestration, crop pollination, managed timber production, water purification (for nutrients), reservoir hydropower production, and sediment retention for reservoir maintenance. It also includes a biodiversity model so that tradeoffs between biodiversity and ecosystem services can be assessed. Models are under development for other services. These include models for flood mitigation, irrigation, agricultural production, open access harvest of timber and non-timber products, recreation and tourism, and cultural benefits.

To date, the marine and terrestrial/freshwater models are treated separately. InVEST 2.0 and this version of the guide presents the models for the two systems in turn. However, in future releases, models for the two systems will be more integrated. This will occur in two primary ways. First, some models will have the flexibility to be applied in either terrestrial or marine systems (e.g. carbon storage and sequestration, biodiversity, tourism and recreation, aesthetic views). (In fact, the terrestrial biodiversity model can be applied, as is, to nearshore marine systems). Second, we are working to link freshwater and marine models so that effects of watershed activities on coastal and marine systems can be explored. Such linkages will be included in later releases of InVEST (likely 2012).

InVEST is most effectively used within a decision-making process that starts with a series of stakeholder consultations (illustrated in Figure 1). 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.

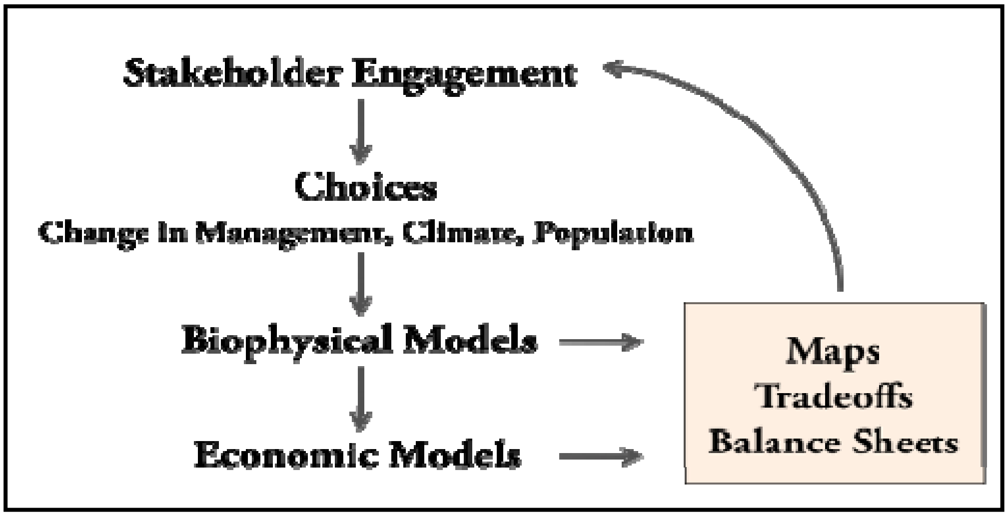


***Figure 1**. Schematic of the decision-making process in which InVEST is embedded. Stakeholders create scenarios that are assessed for ecosystem service value by biophysical and economic models that produce several types of outputs.*

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* (e.g., tons of carbon sequestered) or *economic terms* (e.g., net present value of that sequestered carbon). The spatial resolution 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 in an iterative process, these stakeholders may choose to create new scenarios based on the information revealed by the models until suitable solutions for management action are identified.

InVEST has a **tiered** design. **Tier 0** models map relative levels of ecosystem services and/or highlight regions where particular services are in high demand. For example, the coastal exposure model in InVEST 2.0 maps regions of the coastline that are particularly susceptible to erosion and flooding. It does not use a production function to yield outputs of meters of shoreline eroded or to value coastal protection services provided by nearshore marine habitats. There is no valuation done in tier 0 models. **Tier 1** models are theoretically grounded but simple. They are the simplest models we could create that capture the essence of the key processes in question. They are suitable when more data are available than are required for Tier 0, but they still have relatively simple data requirements. Tier 1 models can identify areas of high or low ecosystem service production and biodiversity across the landscape, and the tradeoffs and synergies among services under current or future conditions. All tier 1 models give outputs in absolute terms, and provide the option for economic valuation (except for biodiversity). Continuing the coastal protection example, the Tier 1 coastal protection model (to be released in May 2011), uses a production function to yield biophysical (meters not eroded or area of flooding avoided), economic (property damage avoided) and social (number of people protected) outputs that can be attributed to nearshore vegetation.

More complex **Tier 2** models are under development for biodiversity and some ecosystem services. Tier 2 models provide increasingly precise estimates of ecosystem services and values, which are especially important for establishing contracts for payments for ecosystem services programs or assessing scenarios that address change on a sub-annual basis. For example, scenarios that represent a change in the monthly or seasonal timing of fertilizer application or water extraction in agricultural systems cannot be assessed by Tier 1 models, but will be treated well by Tier 2 models. It is expected that users will be able to mix and match Tier 0, 1, and 2 models to create the best suite of models given past work, existing data, and the questions of interest. Although the more sophisticated models require substantial data and time to develop, once they are parameterized for a certain location, they will provide the best available science for new decisions. In some cases (e.g. for fisheries), complex tier 3-type models already exist in a particular location. The Natural Capital Project will not develop new tier 3 models, but rather sees these as the sophisticated, dynamic models usually developed for individual sites or contexts. We aim to develop the capability of InVEST to communicate with such existing, complex models so that InVEST inputs (e.g. scenarios) can be fed in, and outputs from those complex models can be compared with other InVEST outputs.

Most of the models in InVEST 2.0 are tier 1 models (though there are some tier 0). Tier 2 models for several services have been formulated and documented in a book in press with Oxford University Press (available April 8, 2011)). We will design the Tier 2 software platform as a space where Tier 0, 1, 2 and 3 models can be integrated as appropriate for different applications.

TIER 0 Models	TIER 1 Models	TIER 2 Models	TIER 3 Models
Relative values	Absolute values	Absolute values	Absolute values
No valuation	Valuation done through a suite of approaches	Valuation done through a suite of approaches	Valuation done through a suite of approaches
Generally not time-specific, or annual average	Annual average time step, no temporal dynamics	Daily to monthly time step, some temporal dynamics	Daily to monthly time step, temporal dynamics with feedbacks and thresholds
Appropriate spatial extent ranges from sub-watershed to global	Appropriate spatial extent ranges from sub-watershed to global	Appropriate spatial extent ranges from parcel to global	Appropriate spatial extent ranges from parcel to global
Good for identifying key areas (relatively high risk or ecosystem service provision)	Good for strategic decisions with absolute values, can be good for tactical decisions with calibration	Good for tactical decisions with absolute values	More precise estimates of ecosystem service delivery
No ecosystem service interactions	Some ecosystem service interactions	Some ecosystem service interactions	Sophisticated ecosystem service interactions with feedbacks and thresholds

A work in progress

The development of InVEST is an ongoing effort of the Natural Capital Project. The models included in this Beta release are at different stages of development and testing, however they are all sufficiently developed to be applied. To date, the Beta terrestrial models have been applied in several sites and decision contexts, including to support: policy and conservation planning in the Willamette Basin USA, private landowners in Hawai'i USA, multi-stakeholder planning in Tanzania, permitting and licensing in Colombia, water fund design in Colombia and Ecuador, and priority setting for international aid in the Amazon Basin. Updated and new models for additional ecosystem services will be released as they become available.

This is the first release of InVEST that includes the marine and coastal models. We have made every effort to debug, to test the models for stability, and to make the models easy to use, but we know that improvements will be in order. Please be patient and give us feedback (http://invest.ecoinformatics.org). Also, because only 4 of many models under development are included in this release, we have included brief descriptions of the models we will release in May 2011. (See "What's Coming Up in Future Releases" document at http://invest.ecoinformatics.org).

InVEST is a freely available, open source product and as such the source code of each model can be inspected and modified by users. InVEST is subject to standard open source license and attribution requirements, which are described and must be agreed to in the installation process, As in other open source projects, it is hoped that users will submit improvements, error fixes, and suggestions to the Natural Capital Project so that improvements can be made to future versions.

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.

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).

This guide does not include detailed theoretical discussions of the scientific foundation of each model, which will be published in an upcoming book from Oxford University Press (available April 2011).

GETTING STARTED

Getting Started with InVEST

InVEST tools run as script tools in the ARCGIS ARCTOOLBOX environment. To run InVEST, you must have:

- **ARCGIS 9.3** (service pack 1 or 2). We are working hard to support **ARCGIS** 10, but have not been able to transition all models yet. As of this release, we support ArcGIS 10 for all of the marine and coastal models, timber production, biodiversity, pollination and carbon models.
- **ARCINFO level license** to run some of the models
- **Spatial Analyst extension** installed & activated
- **Python 2.5 or higher**, which is typically installed automatically as part of ARCGIS. The pollination model and all marine models require additional Python libraries.

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

A set of sample data is supplied with the models so you can become familiar with the models and how they work. To use InVEST for your context, however, you must compile the data described in the chapter(s) for the model(s) you wish to run and format them as indicated.

Installing the InVEST tool and data on your computer

The program InVEST-Setup.exe contains the InVEST toolbox, scripts, and training data, and is available at the website http://invest.ecoinformatics.org.

- Double-click on InVEST-Setup.exe to install the contents on your computer. The Setup wizard will prompt you to download the contents at the default path: C:\InVEST\. You can change the location, but make sure the pathname does not have spaces or special characters, as this will prevent the model from working properly. Do not move the folder contents to your desktop as this path contains spaces. While InVEST is not supported for Macs, users who are running a virtualization product (such as Parallels or VMware) should make sure the installation is placed on the C: drive of their virtual machine.

- Using Windows Explorer, take note of the folder structure and files extracted from InVEST-Setup.exe. Within the InVEST folder, you will see the toolbox InVEST20.tbx. The python scripts are in the folder \InVEST\python\. There is one script per model, and each ends with a *.py suffix. In addition, you will see folders for Base Data, Biodiversity, Hydropower, Carbon, Pollination, Sedimentation, Water Purification, and Timber. These folders contain sample data. You will also see an ArcMap file called InVEST20.mxd with the InVEST toolbox pre-loaded.

Downloading and Installing Python Library Extensions

InVEST users running the pollination model in ArcGIS 9.3 or any of the marine models in either ArcGIS 9.3 and 10 are required to download the Python extensions file found on the InVEST installer download page. For the marine models, users can also run the "Marine Python Extension Check" tool to determine which extensions have been properly installed on their machine.

All Marine InVEST tools require 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.
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.

The Terrestrial InVEST Pollination model for ArcGIS 9.3 requires the installation of one extension:

1. Geospatial Data Abstraction Library **(GDAL)** is a translator library for raster geospatial data formats.

*Important note about the NumPy extension: An older version of NumPy comes standard with the ArcGIS 9.3 and 10 installations of Python. 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.

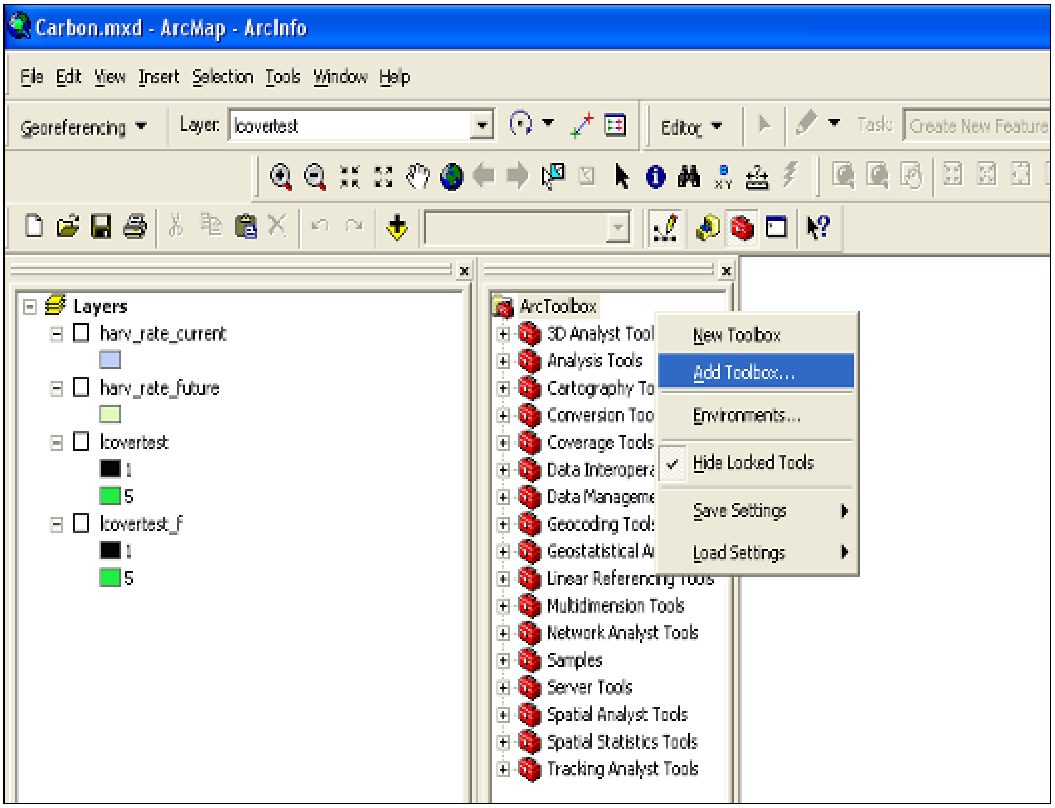
See the Marine InVEST FAQ document (on http://invest.ecoinformatics.org) for help with installing these extensions.

Adding the InVEST toolbox to ArcMap

If you are working with sample data, you may wish to open InVEST20.mxd, which has the toolbox already loaded. Follow these steps if you will be working with your data.

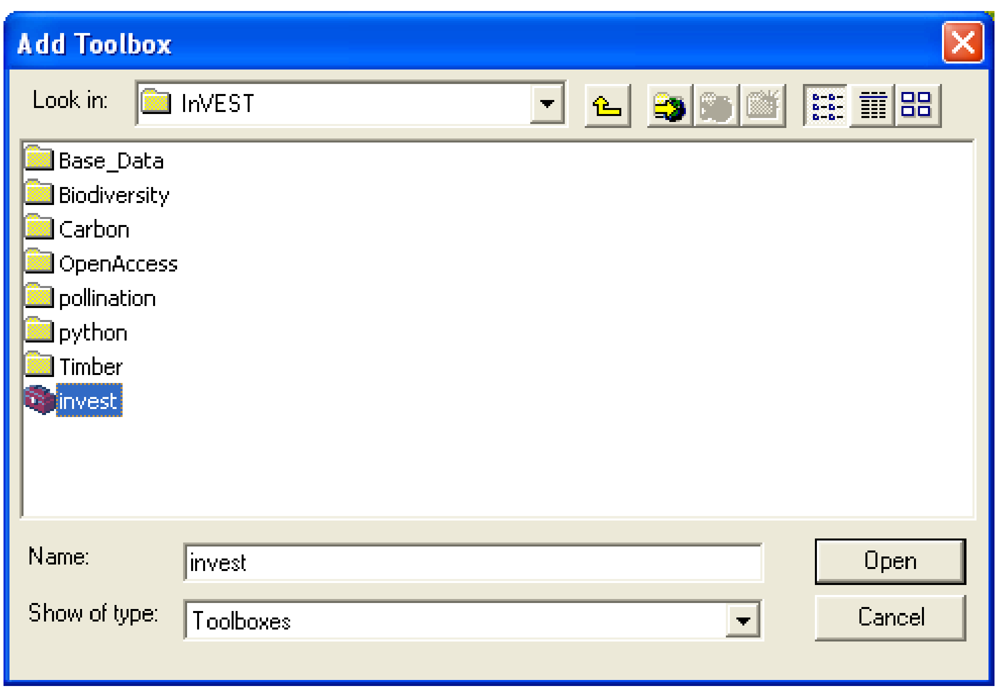
- Start ArcMap. Save as a new mxd file. Ensure that ArcToolbox is open. If not, select the toolbox icon from the standard toolbar.

- Right-click on an empty part of the ArcToolbox window and select Add Toolbox. Or, right click on the top-most ArcToolbox text (see graphic below).



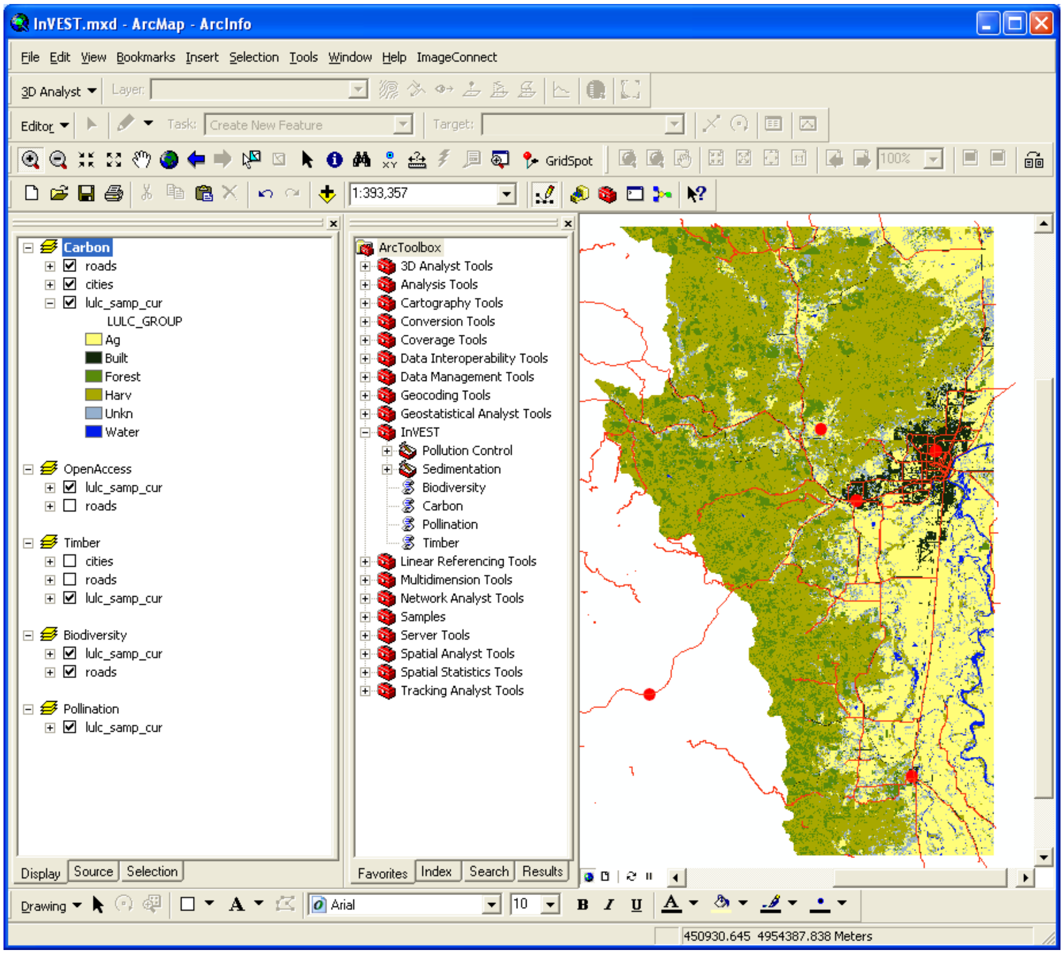
Adding the InVEST toolbox

- Navigate to the location of InVEST20.tbx, in the InVEST folder. Select the toolbox and click OPEN. Do not double click on the toolbox icon.



Select InVEST Toolbox

- The INVEST toolbox should appear in ARCTOOLBOX. Click on the plus sign to the left of INVEST to expand it. You will see scripts for each InVEST model.



InVEST Toolbox and tools displayed

Using Sample Data

The InVEST toolbox comes with sample data, which may be helpful for becoming familiar with the models and 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 outputs respond.

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\pollination\input, and those for the Carbon model in \Invest\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 that are structured like the sample data folders. You will also need to redirect the tool to access your data.

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').
- Raster dataset names cannot be longer than 13 characters and the first character cannot be a number.
- Spatial data should be projected in meters, and all input data for a given tool should be in the same projection. If your data is not projected or it is in a projection that is not in meters, InVEST will warn you and in some cases stop running.
- Depending on the resolution (cell size) of your raster data, the model could take a long time to run. To make the tool run faster, enter a desired resolution that is larger than the original resolution. This will speed up the execution, but will reduce the accuracy of your result. It is recommended to initially run models with large cell sizes to increase speed and reduce memory needs. Final results can be produced with finer resolution.
- Results will be calculated on selections in tables and feature classes. If you are setting the model to read layers and tables from your ARCMAP 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.
- As the models are run, it may be necessary to change values in the input tables. This can happen within ARCMAP 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 ARCMAP, 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. Follow these carefully to ensure your dataset is valid.
- Remember to use the sample datasets as a guide to format your data.

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 and loaded the InVEST toolbox to your ARCMAP document.

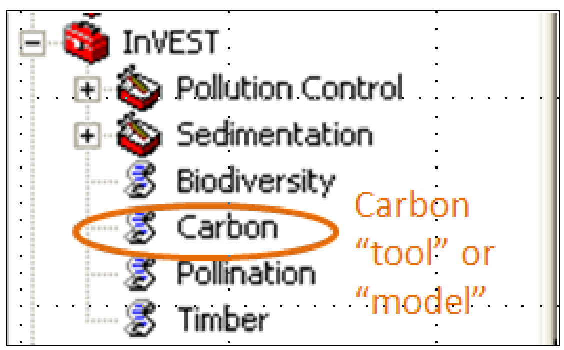
To begin:

- Although not necessary, it's often useful to add your input layers to your ARCMAP document to examine them. Use the ADD DATA button to add input data for a given module.

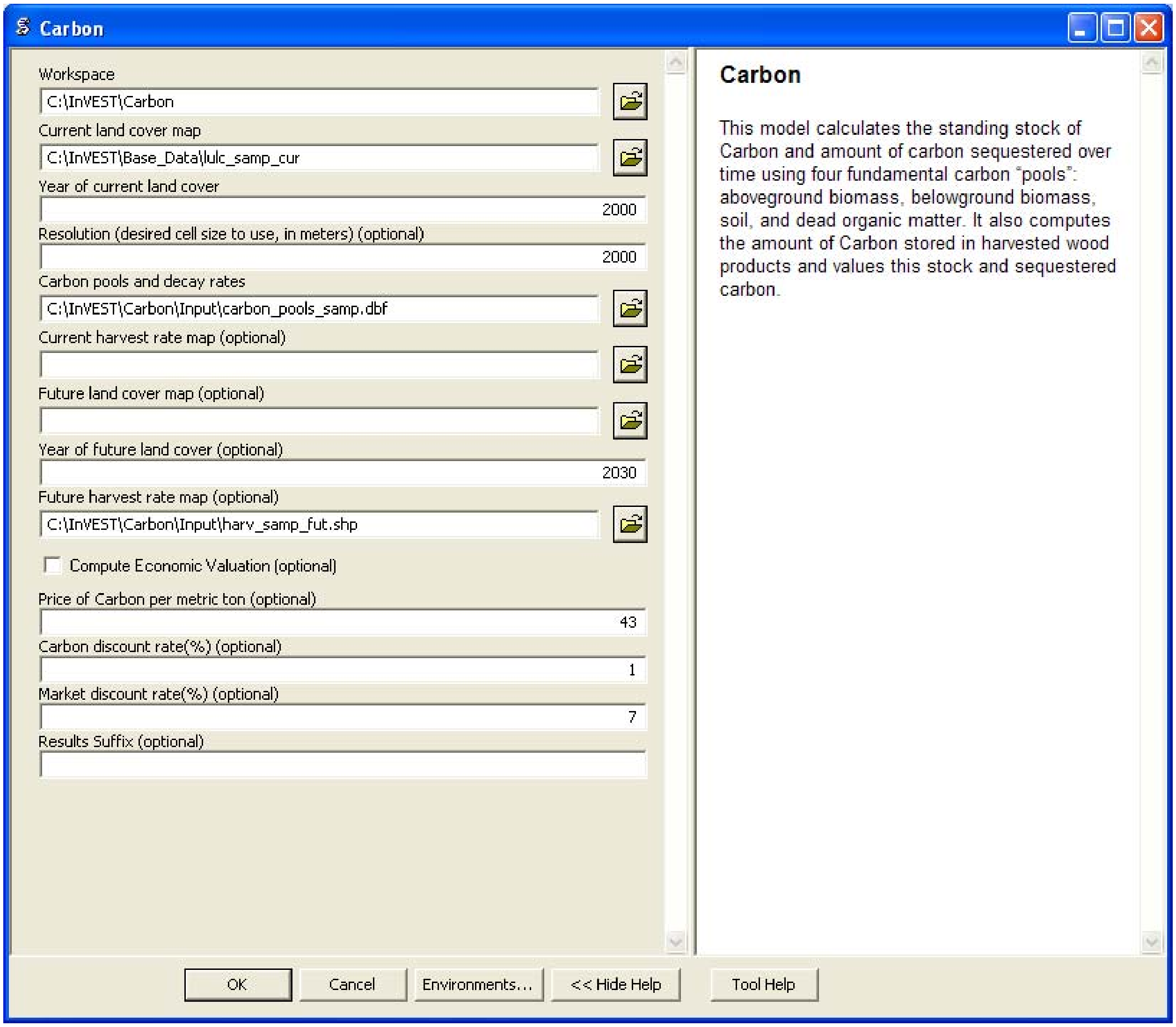
- View the attribute table by right-clicking on the layer and selecting OPEN ATTRIBUTE TABLE. You can change the symbology of an input layer by right-clicking on the layer name in the TABLE OF CONTENTS and selecting PROPERTIES, then clicking on the SYMBOLOGY tab.

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.

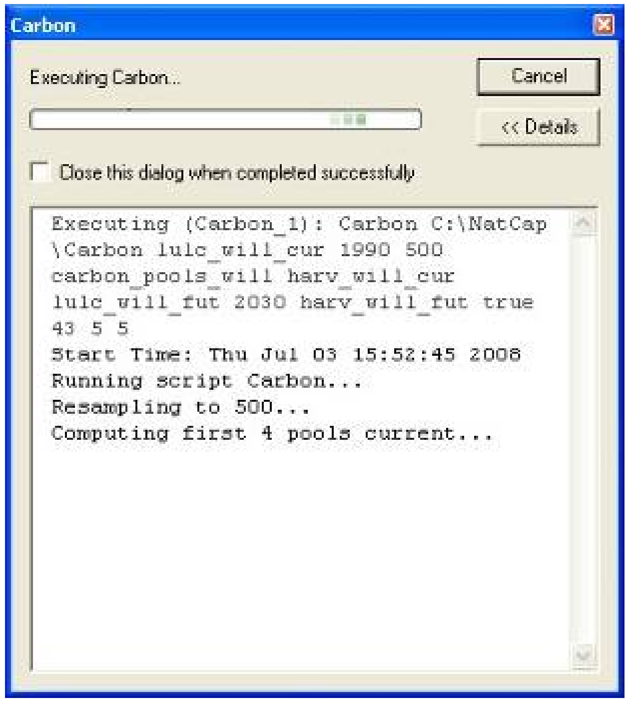


- The Carbon dialog is shown below as an example. Fields for which the entered path leads to a non-existent file will be marked with a red "x" next to the space for that variable. You can run the model with sample data as shown by the default paths, or navigate the paths to your data. Instructions specific 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.



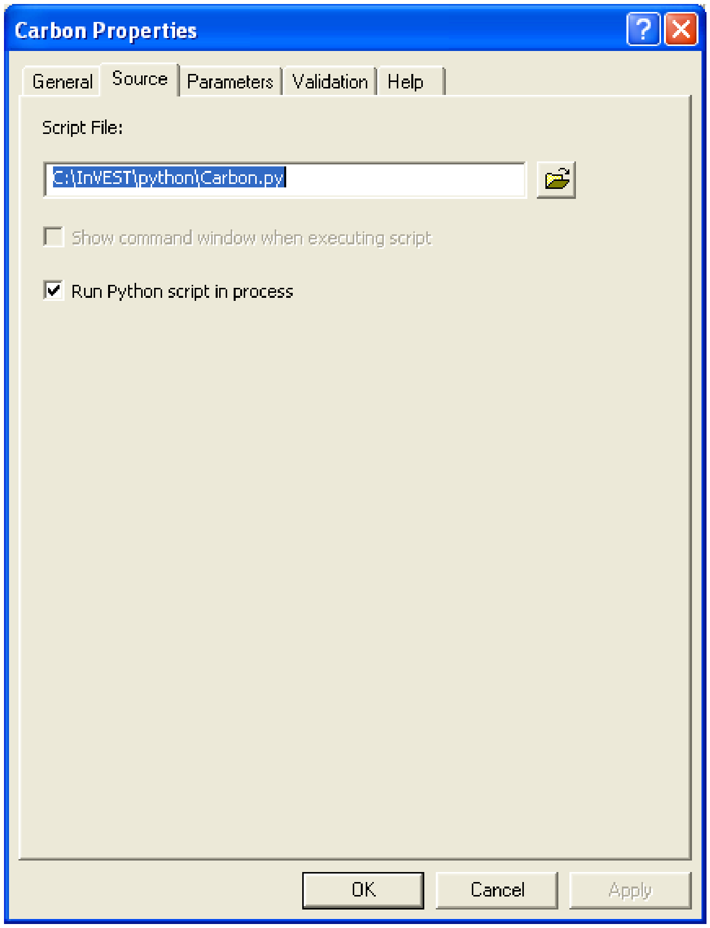
Progress dialog

- 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 ARCMAP 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.

Changing default paths in scripts

If you extracted the files from InVEST-Setup.exe to the default location \InVEST\, the INVEST toolbox will work after you load it to ARCMAP. If you extracted the contents of the folder to a different location, it will work as long as you maintain the internal structure of the InVEST folder. If you moved the python scripts out of the InVEST folder, however, the python scripts associated with the InVEST toolbox still need to be correctly referenced. To do this:

- Add the InVEST toolbox to your ARCMAP document as described above.
- Right-click on the model name in the InVEST toolbox and click on PROPERTIES. For example, click on the plus [+] sign to the left of the InVEST toolbox, and then right-click on the Carbon model. The PROPERTIES dialog appears. Select the SOURCE tab from the top. The path to the python script associated with this tool is shown.



Setting source for the script

- Edit this path to point to the location of the python script. It is easiest to click on the folder button at the right of the box to browse to the script and select it.

- Click OK.

Changing default variables

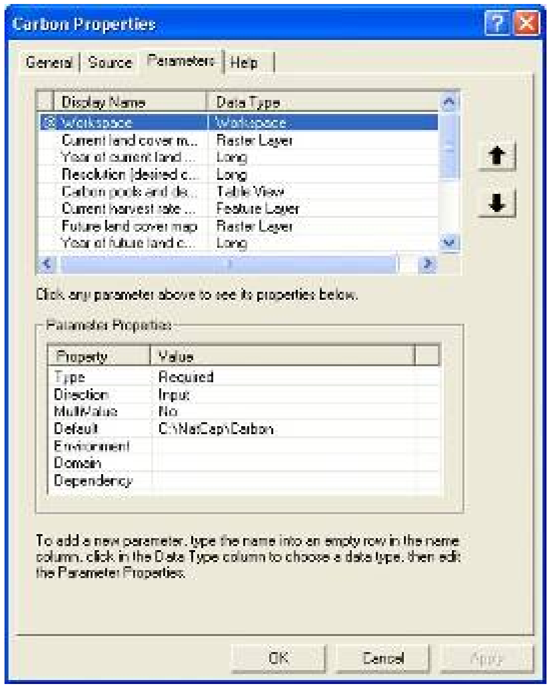
When you double click on an InVEST model such as Carbon or Timber, an interface will appear. These interfaces show default values and path names to sample data (described more fully in the chapters describing each model). You can edit these pathnames to point to data. Another way to change the default path and file names is to:

- Right-click on the model name and click on PROPERTIES in the InVEST toolbox. The dialog below appears.

- Select the PARAMETERS tab at the top to see a list of input parameters for the model. Select an input parameter from the top window, which will set its properties on the lower window, including your desired default values and pathnames.

- Click OK after setting the desired defaults for workspace path, and any other defaults.

Important Note: do not change the order or data type of parameters in the top window, since the program calls these in order. Changes to the order or data type will cause the script to fail.



Setting parameter properties

- Click OK when you have set your desired defaults for workspace path, and any other defaults.

Support Information

Authorized users of InVEST (i.e., those who have obtained the software by registering and receiving a password to download it) have access to limited online support at http://invest.ecoinformatics.org. Users can submit questions, formal error reports, bug fixes, or modified versions of the code to contribute to the next version of the open source product. You must register to receive support and access to the user community.

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 (www.naturalcapitalproject.org). This site is also a good source of general information on InVEST and other activities of the Natural Capital Project.

Model run checklist

Use this checklist to ensure that the models run successfully.

- ☐ ArcGIS Version: As stated above not all ArcGIS versions are supported. Most models are tested in ArcGIS 9.3 SP2 or ArcGIS 10 (for the selected models currently supported). It is advisable to upgrade to one of these versions.
- ☐ Python Extensions: For all marine models, ensure that the latest Python library extensions have been installed: 1) NumPy, 2) SciPy, and 3) PythonWin. Additionally, Microsoft Excel is required to run the marine models. For ArcGIS 9.3 users, the pollination model requires installation of the GDAL library.
- ☐ Spatial Analyst extension: Most of the models require ArcGIS spatial analyst extension. Ensure that this is installed.
- ☐ 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.
- ☐ Folder naming: ArcGIS is strict about folder naming. Avoid spaces and special characters in file and folder names.

Reporting errors

If you experience errors running the models you can get assistance from the discussion list mentioned above. Provide the following details in order to get quick help:

1. The model in which you encountered the error.
2. Your ArcGIS version and service pack.
3. The error text (copy and paste this from the tool dialog, including all the progress report in the tool dialog. Note that the right click does not work in the dialog so use Ctrl+C to copy the error).
4. Indicate whether you were running with sample data or your own data. Ensure you can successfully run with sample data before you try with your own data. This confirms that your system is well setup and ready to run the models.
5. It is preferable to include the parameter file. The models output a parameters file that indicates your input parameters. This can be helpful in troubleshooting.
6. Make a distinction between errors and features missing from the model. If the issue you are facing is related to the model design, please give a clear explanation of this so that the model-development lead will be able to review the issue and provide support.

Working with the DEM

For the hydrology tools Water Purification: Nutrient Retention and Avoided Reservoir Sedimentation, having a well-prepared digital elevation model (DEM) is critical. It must have no missing data or circular flow paths 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 ArcMap Spatial Analyst functions as well as

ArcHydro(see resources below), an ArcMap data model that has a more complex and comprehensive set of tools for modeling surface water features. ArcSWAT which is not covered here could be a good easy to use option to delineate and create smaller sub-watersheds. 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.

- The Hydro_layers directory

  When tools are run that use DEM-derived layers like slope and flow direction, the tool looks for a folder called 'Hydro_layers', located in the same folder as the DEM. If this folder does not exist, or any of the required derived layers within the folder don't exist, the tool will generate them from the input DEM, otherwise it uses the layers that already exist. In general, this is convenient and efficient. However, if you decide to use a different DEM than the one that was used to generate the files in Hydro_layers, and the new DEM is located in the same folder as the old DEM, the tool will not realize that it is different, and will continue to use the old derived layers. So in this case it is necessary to delete the Hydro_layers folder before re-running the tool using the new DEM, so that the derived layers are regenerated.

- 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:
  con(isnull([theDEM]), focalmean([theDEM], rectangle, 4, 4), [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

If the stream network generated from the DEM does not correctly match reality, 'burning' a correct stream network into the DEM might be necessary. Here are the basic steps for ArcMap:

1. Create the stream network from the DEM using the Hydrology -> Flow Accumulation tool and compare it to a known correct stream layer. If the generated stream network does not look correct, continue with the following steps.
2. If starting with a vector stream layer, convert it to a grid that has the same cell size and extent as the DEM.
3. Assign the stream grid a cell value of 1 where there are streams and 0 elsewhere.
4. 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 ArcMap, first identify sinks using ArcMap's 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.

- Flow direction loops

If there's a problem in the flow direction raster, such as a loop, the Water Purification and Sedimentation tools may go into an infinite loop and eventually time out, producing this error:

"Error: Sub-watershed 1 is taking too long (45 minutes). This probably indicates that there's a flow direction loop."

Diagnosing and repairing loops is difficult and is beyond the scope of our tools and built-in ArcMap functions. However, a very rough method of determining whether a loop is being encountered is provided in both of the scripts WP_2_Nutrient_Removal.py and Sediment_1_Soil_Loss.py. In each of these files, look for 3 separate commented-out sections of code beginning with 'Flow direction loop debugging'. Uncomment the subsequent lines (containing references to 'outfile') as directed. The next time the tool is run, it will write information to the file
<Workspace>\Output\wp(or sed)_loop_debug_<current time>_<suffix>.txt.
This can become a very large file, as information is recorded on every cell in the watershed raster, as they are processed by moving along flow paths.

Each line of the debug file has three values: the nutrient or sediment load originating on that cell, the flow direction and the fraction of nutrient or sediment retained by that land use class (as given in the input Biophysical table.) With the debugging lines of code uncommented, run the tool. Then look at the end of the debug file - if a loop was encountered, multiple lines with a particular set of values will be repeated. These values can be used to help identify where the loops occur, by retaining the <Workspace>\Intermediate folder (comment out the lines at the bottom of the code under 'Clean up temporary files' before doing the debug run), adding the Intermediate files 'frac_removed_ext', 'flowdir_ext' and 'loads_ext' to the map, and picking out the cells that have the particular set of values that repeated in the debug file (the CON tool can be used for this purpose.) This might produce many different matching areas, which would then have to be further investigated to single out the problem area.

Once a loop is found, it might help to go back to the DEM and do more sink filling, or use the CON tool similarly to how it is used in the "Check for missing data" section above to assign new values.

- Creating watersheds

To create watersheds in ArcMap, 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.

- Creating sub-watersheds

  If the watersheds of interest are too large (greater than 4000 X 4000 pixels) for the Water Purification and Sediment models, they will need to be broken up into sub-watersheds. See their User's Guide sections for more information, under Data Needs/Sub-watersheds.

  In ArcMap the Hydrology -> Watershed tool can be used. 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, such that a sub-watershed will be generated for each.

  Again, after the sub-watersheds are generated, verify that they represent the catchments correctly. Ensure each sub watershed is assigned a unique id and no duplicates are present.

Resources

ArcHydro: http://www.crwr.utexas.edu/giswr/hydro/ArcHOSS/Downloads/index.cfm

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

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: http://support.esri.com.

WAVE ENERGY MODEL

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. This is a "Tier 1" model.

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 simple 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.).

The Model

The objective of the WEM is to help decision-makers and stakeholders inform 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

$$P_n = \frac{\rho g}{16} H_s^2 \cdot C_g(T_e, h) \qquad \text{(Eq. 1)}$$

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}). C_g is a function of wave energy period, T_e (sec), and water depth h (m) (Cornett 2008). C_g can be estimated as

$$C_g = \frac{1}{2}\left[1 + \frac{2kh}{\sinh(2kh)}\right]\sqrt{\frac{g}{k}\tanh(kh)} \qquad \text{(Eq. 2)}$$

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:

$$w^2 = gk\tanh(kh) \qquad \text{(Eq. 3)}$$

An iterative numerical solution scheme can be applied to solve Eq. 3 with initial estimates of $k = \omega^2 / \left(g \cdot \sqrt{\tanh(\omega^2 \cdot h / g)}\right)$. 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 $\alpha = 0.90$ 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 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.

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 for several WEC devices that 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)

Table 2 shows an example of wave energy absorption performances in each seastate bin for Pelamis.

Table 1. Occurrence of hours (hr/yr) in each seastate bin in the west coast of Vancouver Island.

		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

Table 2. Wave energy absorption performance (kW) in each seastate bin for Pelamis.

		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

By multiplying each cell in the annual occurrence of hours table (Table 1) by each corresponding cell of the wave energy absorption performance table (Table 2), 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. As technology progresses, the device-specific parameters will likely need to be updated and new devices added. 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:

$$NPV = \sum_{t=1}^{T} (B_t - C_t)(1+i)^{-t} \qquad \text{(Eq. 4)}$$

and is evaluated over a life span, *T*, of a WEC facility. To discount the value of future benefits and costs, we use a default discount rate, *i*, of 5 percent. 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.

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 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 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.

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.

[1] Both the discount rate and the wholesale price of electricity are user-defined inputs. We provide a default value of 5% for the discount rate and .20 cents for the wholesale price of electricity. 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 converted to 2009 $USD.

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.

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.

A. 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 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 “WaveEnergy” 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\WaveEnergy\runBC

2. **Path to Folder with Wave Base Data (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.

[3] Wallace and Dunnett (2009) model 24 devices in their application.

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.

   Preset Areas:
   - West Coast of North America and Hawaii
   - East Coast of North America and Puerto Rico
   - Global (Eastern Hemisphere)
   - 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.

   Table Names: File can be named anything, but no spaces in the name
   File type: *.xls or .xlsx (if user has MS Excel 2007 or newer)
   First row: wave period bins (T_p) in second
   First column: wave height bins (H_s) in meter
   The numbers in the table indicates captured wave energy for the given seastate condition defined by wave height (H_s) and period (T_p).
   Sample data set:
   \InVEST\WaveEnergy\Input\Machine_Pelamis.xls\Pelamis_performance$

Hs(m) /Tp(sec)	0.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0
0.0	0	0	0	0	0	0	0	0	0	0	0
0.1	0	0	0	0	0	0	0	0	0	0	0
0.5	0	0	0	0	0	0	0	0	0	0	0
1.0	0	0	0	11	27	50	62	64	57	49	41
1.5	0	0	0	26	62	112	141	143	129	110	91
2.0	0	0	0	66	109	199	219	225	205	195	162
2.5	0	0	7	93	171	279	342	351	320	274	230
3.0	0	0	91	180	246	402	424	417	369	343	331
3.5	0	0	86	211	326	484	577	568	502	421	394
4.0	0	105	216	326	394	632	616	583	585	494	454

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: *.xls or .xlsx (if user has MS Excel 2007 or newer)
Sample data set: \InVEST\WaveEnergy\Input\Machine_Pelamis.xls\Pelamis_parameter$

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.

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 valuation to determine the cost to send mooring cables to the ocean floor before running them to landing 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 1 arc-minute resolution, we recommend using the global DEM 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

B. Optional Inputs

The next series of inputs are optional, but may be required depending on other decision inputs.

7. **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 a polygon feature. 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

8. **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 #7). Additionally, the following inputs (#9-12) must be also be specified in order to output economic analysis.

9. **Landing and Power Grid Connection Point Table *(optional, but required for economic valuation)*.** When running the economic analysis, you must provide an Excel spreadsheet 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. Currently, the model allows for multiple landing points, but only one grid-connection point. The model will use this input to create a point feature class and project it based on the projection file specified in input #12. 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. The input is not case sensitive, but does require exact wordings to differentiate the two types.

   Table Names: File can be named anything, but no spaces in the name
   File type: *.xls or .xlsx (if user has MS Excel 2007 or newer)
   Sample data set: \InVEST\WaveEnergy\Input\LandGridPts_WCVI.xls\WCVI$

ID	LAT	LONG	TYPE	NAME	NOTE
1	48.92100	-125.54200	LAND	Ucluelet	Underwater Cable Landing Point
2	49.13900	-125.91500	LAND	Tofino	Underwater Cable Landing Point
3	48.99700	-125.58300	GRID	Ucluelet	Power Grid Connection Point

10. **Economic Parameter 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.

   Table Names: File can be named anything, but no spaces in the name
   File type: *.xls or .xlsx (if user has MS Excel 2007 or newer)
   Sample data set: \InVEST\WaveEnergy\Input\Machine_Pelamis.xls\Pelamis_econ$

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.

11. **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).

12. **Projection *(optional, but required for economic valuation)*.** The model uses this input projection file to accurately project the wave points (contained within the folder from input #2) into a projection with meters as the units. Initially, the input points are unprojected (Geographic - WGS84). In order to accurately calculate the distance and resulting cable costs for wave machine facility sites to land, the model must project all facility site points within the clipped AOI extent. Additionally, so that the model does not have to anticipate datum transformations, the projection file must have a WGS84 datum. At the start, the model will check that this projection input meets these criteria. If not, it will stop and provide feedback.

    <u>File type:</u> projection files provided by ArcGIS (.prj)
    <u>Sample path</u>: Coordinate Systems\Projected Coordinate Systems\UTM\WGS 1984\WGS 1984 UTM Zone 10N.prj

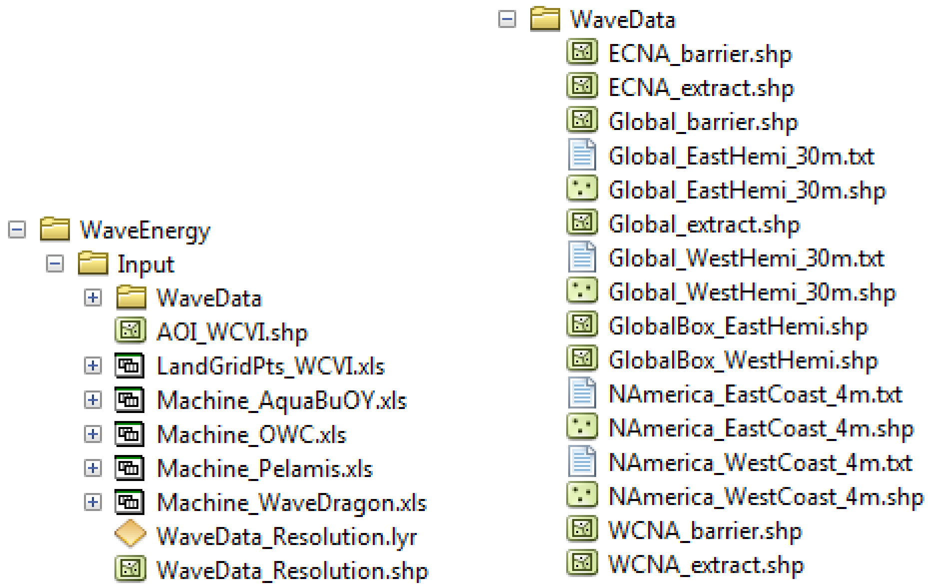
Running the Model

<u>Note about terminology used here</u>: 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.

1. **Exploring the workspace and input folders.** These folders will hold all input, intermediate 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.

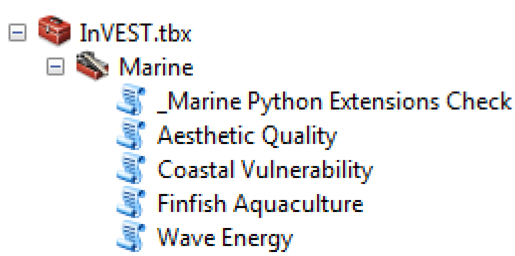
   <u>Exploring a project workspace and Input data folder</u>. The *\InVEST\WaveEnergy* folder holds the main working folder for the model and all other associated folders. Within the *WaveEnergy* folder there will be a subfolder named *'Input'*. This folder holds most of the GIS and tabular data needed to setup and run the model.

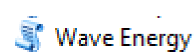
   The following image shows the sample folder structure and accompanying GIS data. We recommend using this folder structure as a guide to organize your workspaces and data. Refer to the screenshots below for examples of folder structure and data organization.



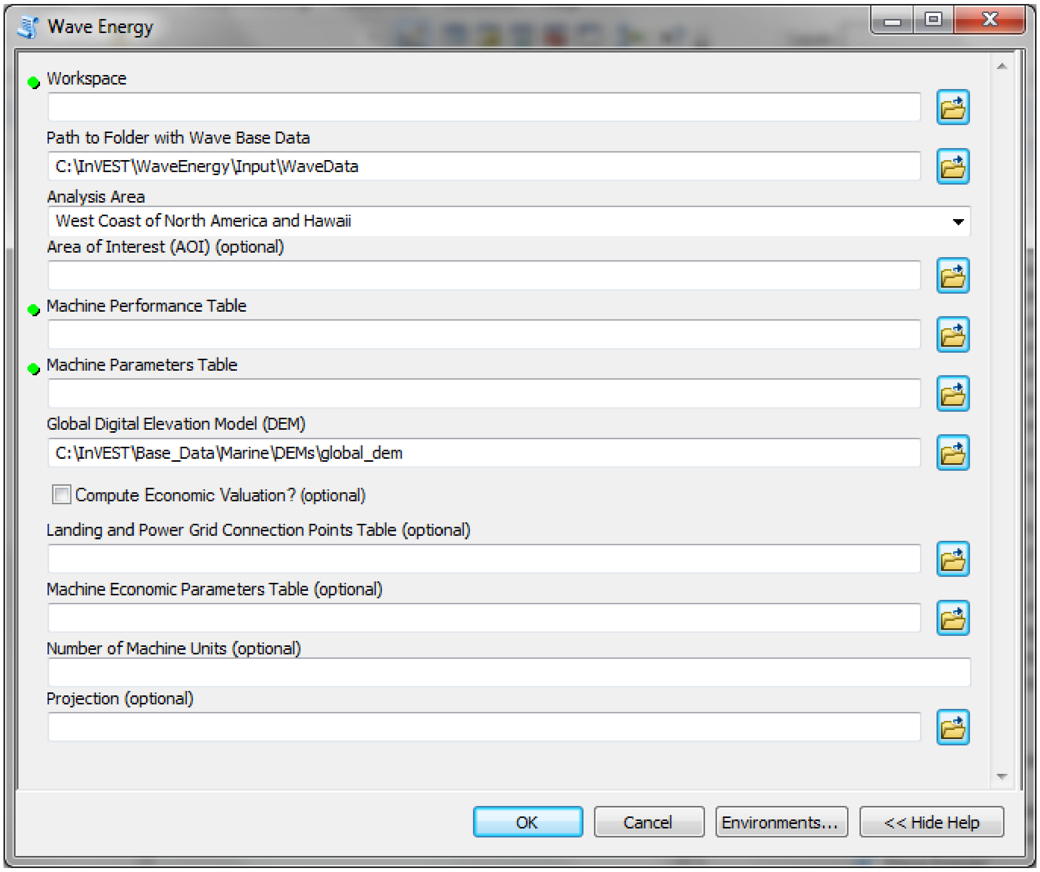
2. **Creating a run of the model.** The following example describes how to set up the Wave Energy model using the sample data provided with the InVEST download. We expect users to have location-specific data to use in place of the sample data. These instructions provide only a guideline on how to specify to ArcGIS the various types of data needed and do not represent any site-specific model parameters. See the **Data Needs** section for a more complete description of the data specified below.

    a. Click the plus symbol  next to the InVEST toolbox.

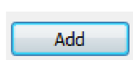
    

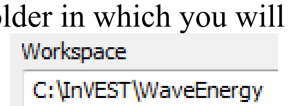
    b. Expand the Marine toolset and click on the Wave Energy script 

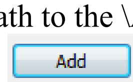
    to open the model.



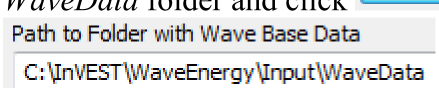
c. <u>Specify the Workspace</u>. Open  the *InVEST* workspace. If you created your own workspace folder (Step 1), then select it here.

Select the *WaveEnergy* folder and click 

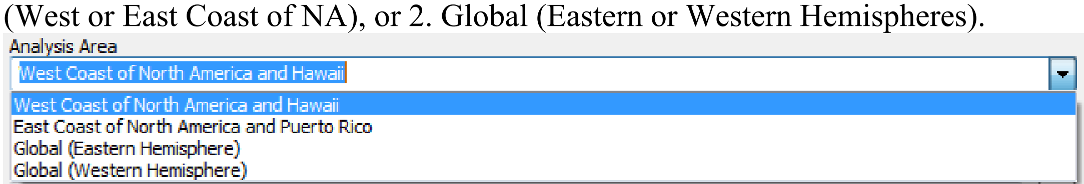
to set the main model workspace. This is the folder in which you will find the intermediate and final outputs when the model is run. 

d. <u>Specify the Folder with Wave Base Data</u>. The model requires the folder location of the wave data. Click  and path to the *\InVEST\WaveEnergy\Input* folder. Select the *WaveData* folder and click 

to set the wave data folder.



e. <u>Specify the Analysis Area</u>. You can run the model at one of two scales: 1. Regional (West or East Coast of NA), or 2. Global (Eastern or Western Hemispheres).



f. Specify the Area of Interest (AOI). The model does not require an AOI, unless the user chooses to run the economic valuation. However, the AOI does permit the user to perform more local analysis if the analysis area (specified above) is too large. This example refers to the *AOI_WCVI.shp* shapefile supplied in the sample data. You can create an AOI shapefile by following the **Creating an AOI** instructions in the FAQ document at http://invest.ecoinformatics.org.
Click and path to the *\InVEST\WaveEnergy\Input* data folder.

If you created your own *Input* folder in step 1b, then select it here. Select the *AOI_WCVI.shp* shapefile and click Add to make the selection.

Area of Interest (AOI) (optional)

C:\InVEST\WaveEnergy\Input\AOI_WCVI.shp

g. Specify the Machine Performance Table. The model requires an Excel table of machine performance characteristics. Click and path to the *\InVEST\WaveEnergy\Input* data folder. Double left-click *Machine_AquaBuOY.xls* and select the worksheet *AquaBuOY_performance$*.

Click Add to make the selection.

Machine Performance Table

C:\InVEST\WaveEnergy\Input\Machine_AquaBuOY.xls\AquaBuOY_performance$

Note: ArcGIS and the model may not recognize the Excel sheet as valid data if it is added to the ArcMap Data View. It is best to add Excel data directly to the model using the Open and Add buttons and navigating to the data.

h. Specify the Machine Parameters Table. The model requires an Excel table of the physical specifications for a specific type of wave machine. Click and path to the *\InVEST\WaveEnergy\Input* data folder. Double left-click *Machine_AquaBuOY.xls* and select *AquaBuOY_parameter$*. Click Add to make the selection.

Machine Parameters Table

C:\InVEST\WaveEnergy\Input\Machine_AquaBuOY.xls\AquaBuOY_parameter$

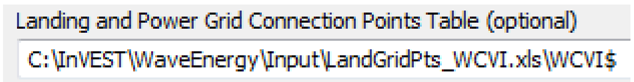
i. Specify the Digital Elevation Model. The digital elevation model provides the base data for the Wave Energy model. Click and path to the *\InVEST\BaseData\Marine\DEMs* data folder. Select the *global_dem* raster, click Add to make the selection and add it to the Wave Energy model dialog window.

Global Digital Elevation Model (DEM)

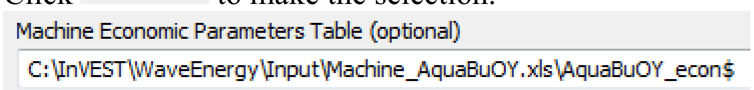
C:\InVEST\Base_Data\Marine\DEMs\global_dem

j. Specify the Economic Valuation (Optional). To conduct economic valuation of the wave energy conversion machines, click the checkbox. Economic analysis is only available if an AOI was specified. ☐ Compute Economic Valuation? (optional)

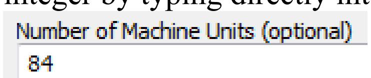
k. Specify the Landing and Grid Points Table (Optional). To conduct the economic analysis the model requires an Excel table of machine locations. Click and path to the *\InVEST\WaveEnergy\Input* data folder. Double left-click *WCVI_LandGridPts.xls* and select *WCVI$*. Click Add to make the selection.



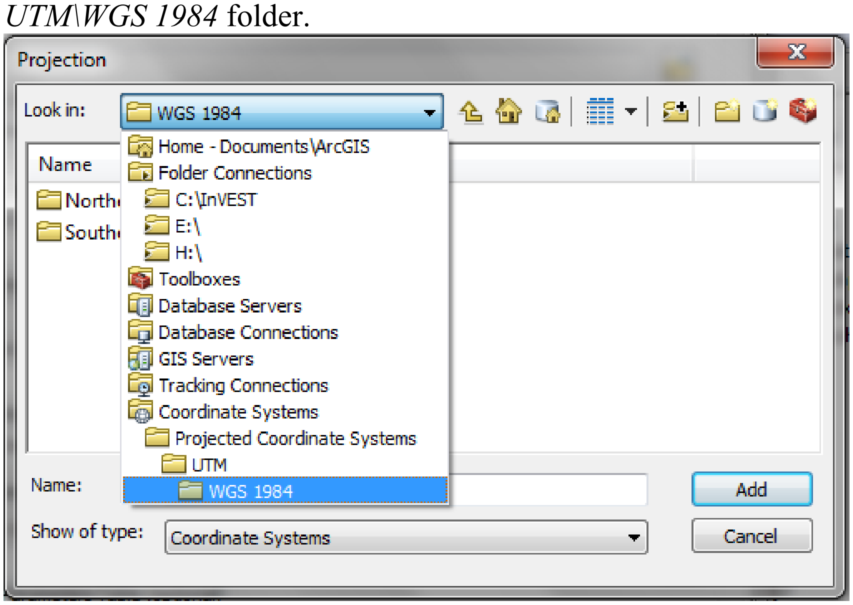
l. Specify the Machine Economic Parameters Table (Optional). To conduct the economic analysis the model requires a table of economic valuation parameters. Click and path to the *\InVEST\WaveEnergy\Input* data folder. Double left-click *Machine_AquaBuOY.xls* and select *AquaBuOY_econ$*. Make sure you select the worksheet that corresponds to the correct wave machine specified in Steps 2g and 2h. Click Add to make the selection.

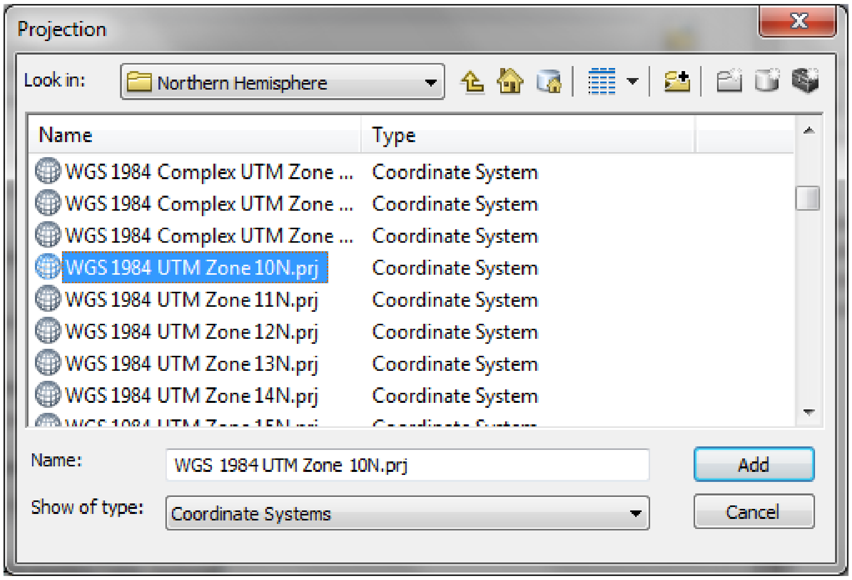


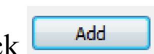
m. Specify the Number of Machine Units (Optional). The model requires the number of machines to perform the economic valuation. Enter the number of machines as an integer by typing directly into the text box.



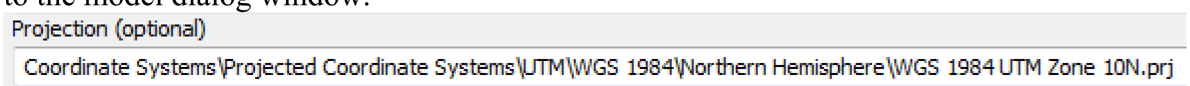
n. Specify the Projection file. The Projection file is specified to set the projection and coordinate information necessary to run the economic valuation. Open the Coordinate Systems folder near the bottom of the Look In list and path to the *UTM\WGS 1984* folder.





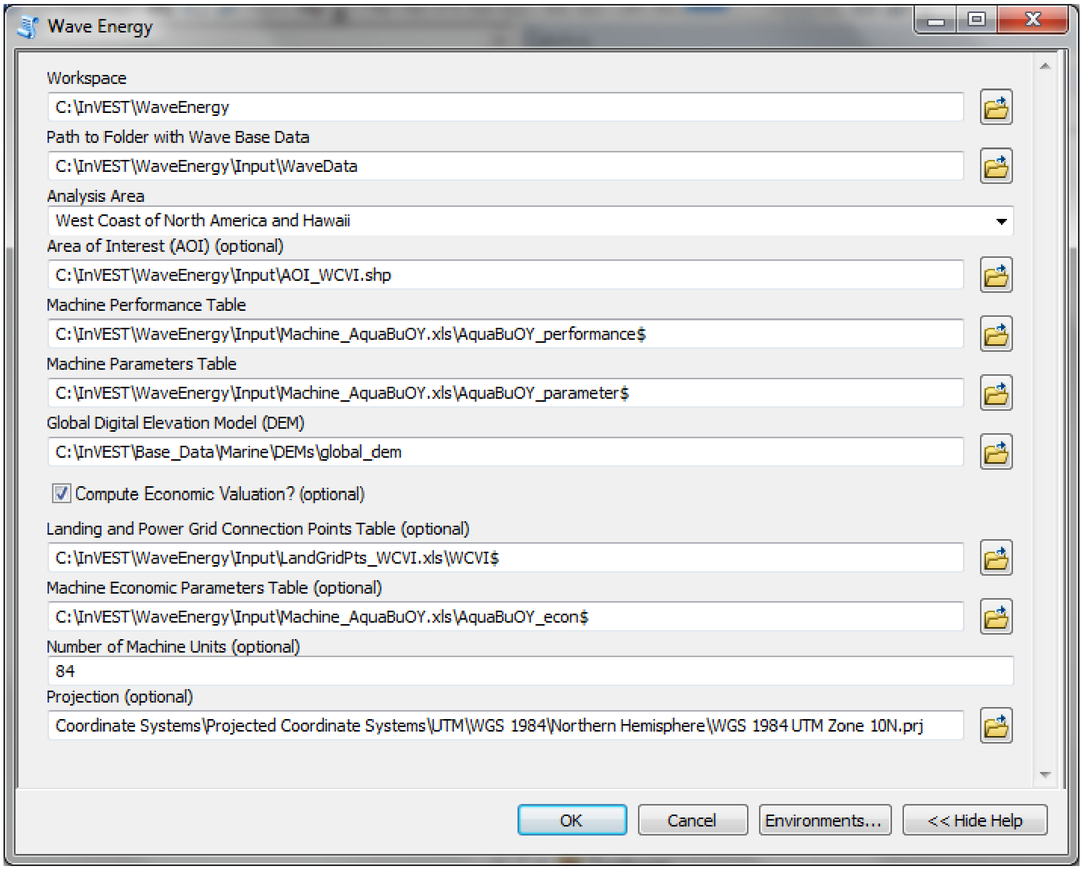
Select the WGS 1984 UTM Zone 10N.prj projection file and click 

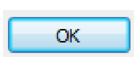
to add it to the model dialog window.



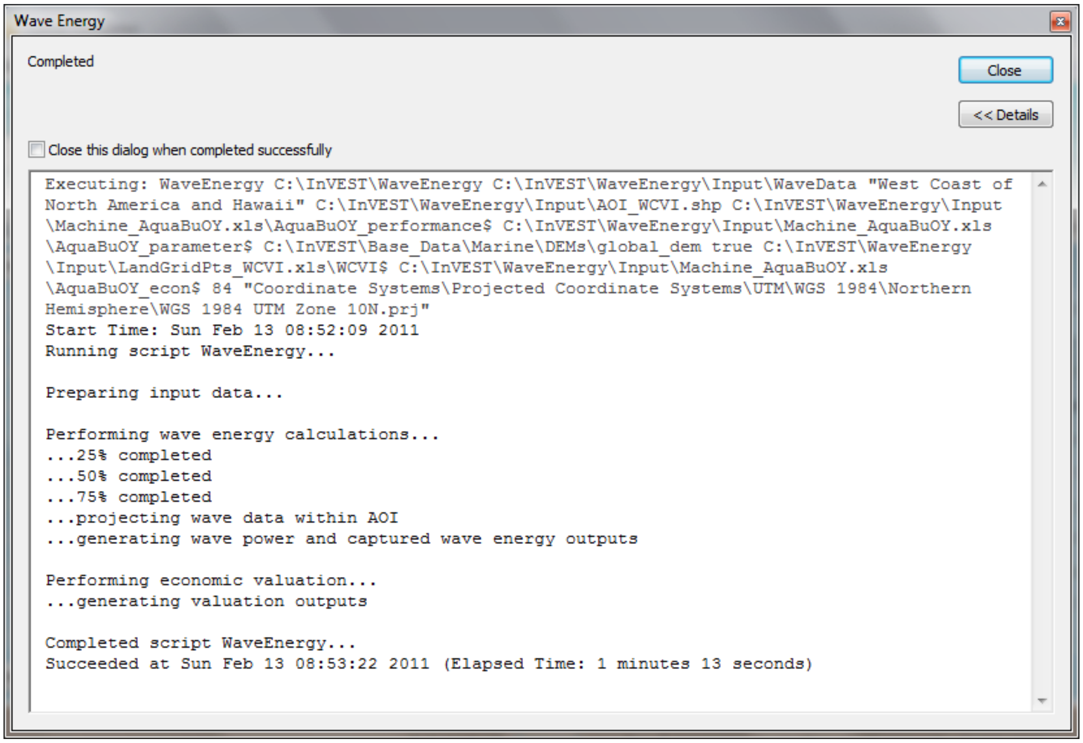
Note: It is assumed that all of your input data are in the same projection and coordinate systems with matching datum. If you need to (re-)project your data, see the Projection section in the Marine InVEST FAQ document found at http://invest.ecoinformatics.org.

o. At this point the model dialog box is completed for a complete run of the Wave Energy model.



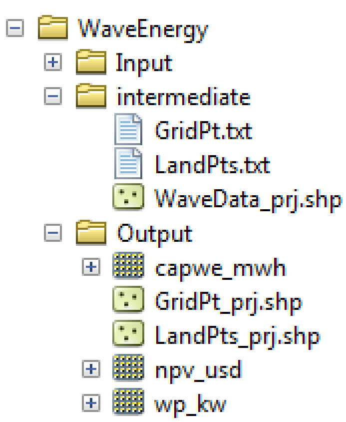
Click 

to start the model run. The model will begin to run and will 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 necessary for the model run.



3. **Viewing output from the model.**

Upon successful completion of the model run, you will see new folders in your Workspace called "*intermediate*" and "*Output*". The *Output* folder, in particular, may contain several types of spatial data, which are described in the **Interpreting Results** section of this guide.



You can view the output spatial data in ArcMap (from either the Intermediate or Output folders) 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".

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:

<u>"Output" Folder</u>

1. Output\ wp_kw
    - This raster layer depicts potential wave power in kW/m for the user-specified extent.
    - 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.
2. Output\ capwe_mwh
    - This raster layer depicts captured wave energy in MWh/yr per WEC device for the user-specified extent.
    - The captured wave energy map provides useful information to compare the performance of different WEC devices as a function of site-specific wave conditions.
3. Output\npv_usd
    - This raster layer depicts net present value in thousands of $ over the 25 year life-span of a WEC facility for the user-specified extent.
    - 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.
    - This is only an output if you have chosen to run economic valuation.

4. Output\ LandPts_prj.shp and GridPt_prj.shp
    - These feature layers contain information on underwater cable landing location and power grid connection points, which have been projected based on the projection specified (input #12) and the coordinates specified in the Excel table for input #10.
    - 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.

5. 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.

<u>"Intermediate" Folder</u>

1. intermediate\WaveData_prj.shp *or* WaveData_clipZ.shp (depending on whether economic valuation is conducted)
    - 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. "I" and "J" – index values for the wave input grid points
        ii. "LONG" and "LAT" – longitude and latitude of the grid points
        iii. "HSAVG_M" – wave height average [m]
        iv. "TPAVG_S" – wave period average [second]
        v. "DEPTH_M" – depth [m]
        vi. "WE_KWM" – potential wave power [kW/m]
        vii. "CAPWE_MWHY" – captured wave energy [MWh/yr/WEC device]
        viii. "W2L_MDIST" – Euclidean distance to the nearest landing connection point [m]
        ix. "LAND_ID" – ID of the closest landing connection point
        x. "L2G_MDIST" – Euclidean distance from "LAND_ID" to the nearest power grid connection point [m]
        xi. "UNITS" – number of WEC devices assumed to be at this WEC facility site
        xii. "CAPWE_ALL" – total captured wave energy for all machines at site [MWh/yr/WEC facility]
        xiii. "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.

2. intermediate\GridPt.txt *and* LandPts.txt
    - These text files log records of the grid and landing point coordinates specified in the Excel table for input #9
    - This is only an intermediate output if you choose to run economic valuation.

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 Grille 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.

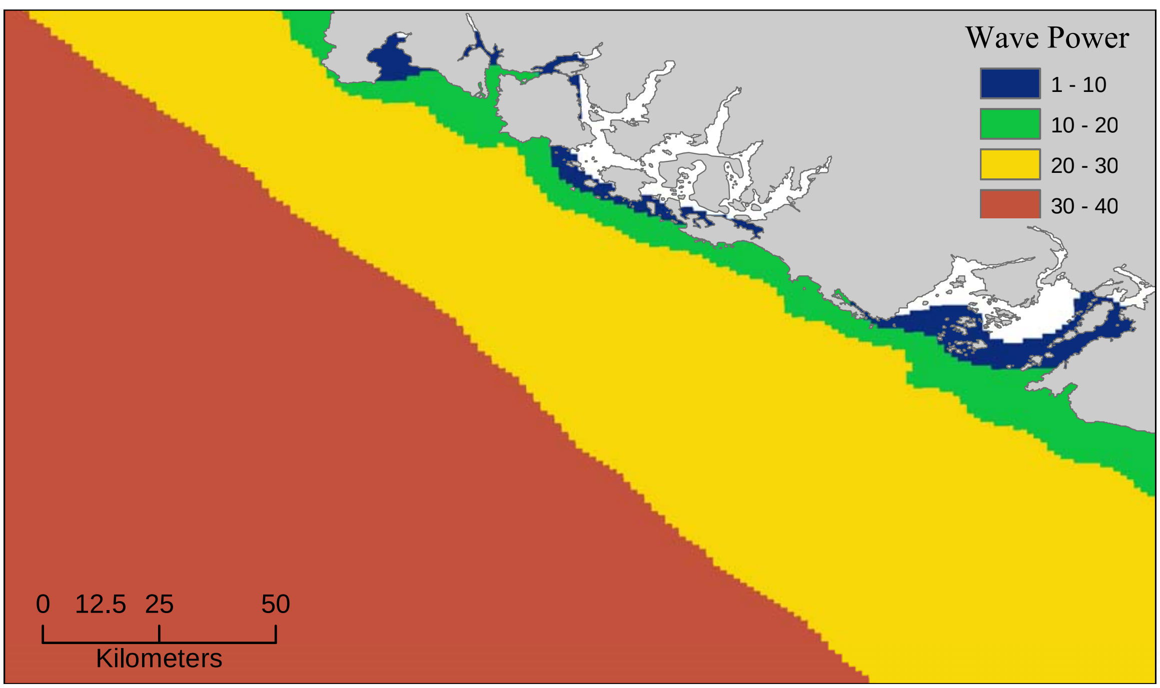


Figure 1. Wave power potential (kW/m) in the west coast of Vancouver Island.

Captured wave energy in this example is calculated based on Pelamis devices with 750 kW power rating (Figure 2). 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.

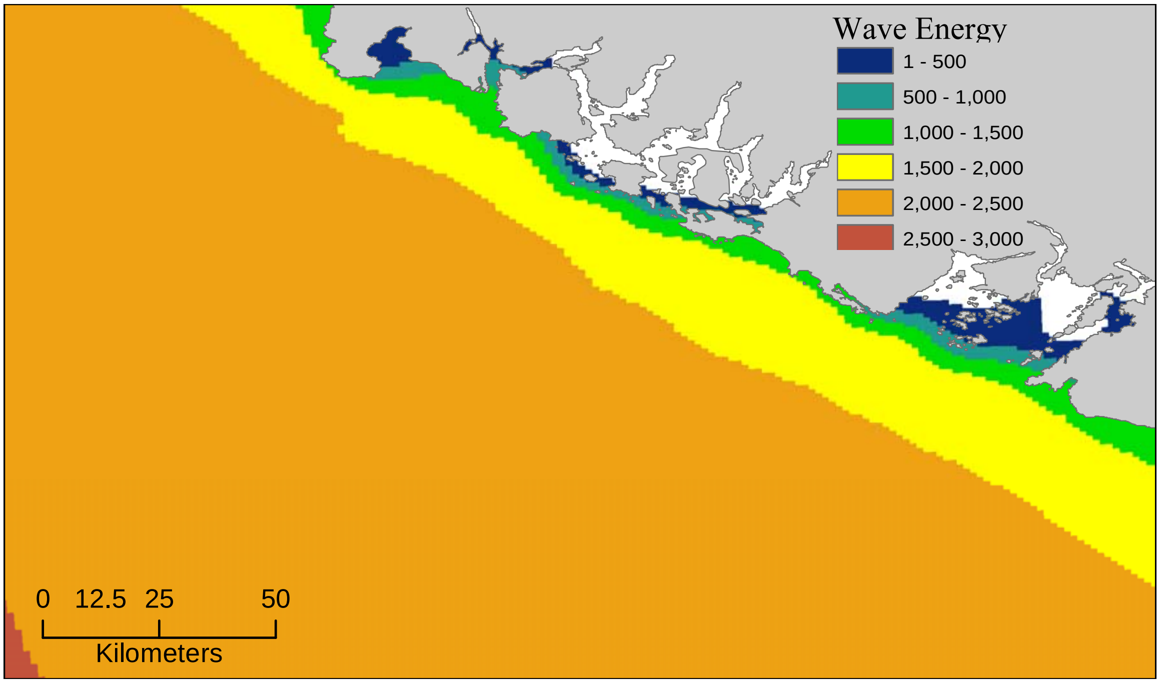


Figure 2. Captured wave energy (MWh/yr) using a Pelamis device with a 750 kW power rating.

For the economic valuation of harvested wave energy, we calculate and map NPV over the 25-yr life-span of a WEC facility (Figure 3). 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.

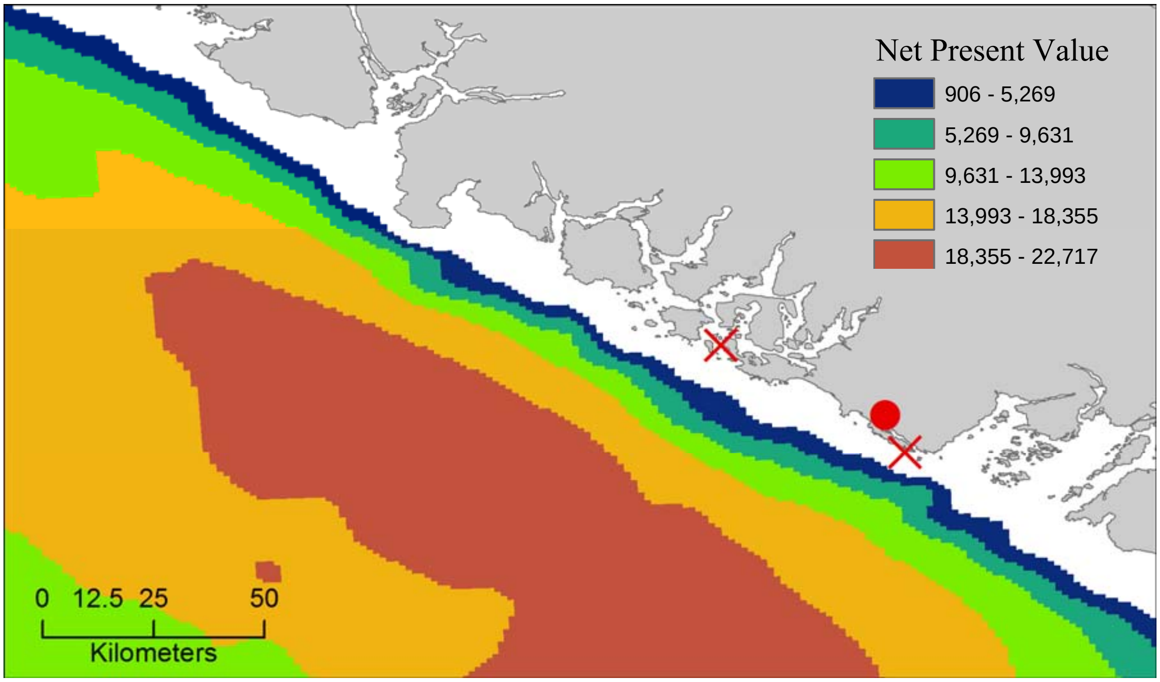


Figure 3. Net present value (thousand $) over a 25-yr life-span, using $100,000 per km for the cost of underwater transmission cables. Two underwater cable landing points are located in Tofino and Ucluelet (×) and the power grid connection point is located in Ucluelet (●). Each of the WEC facilities 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 in Figure 3 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 (Figure 4).

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.

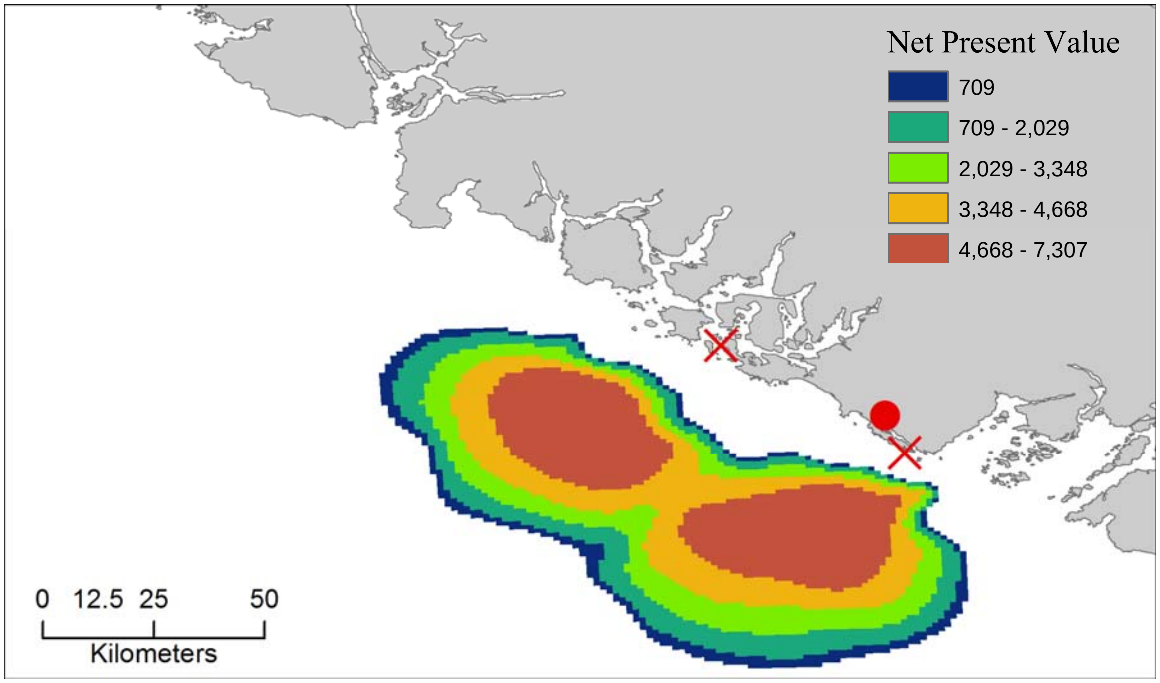


Figure 4. Net present value (thousand $) over a 25-yr life-span, using $ 500,000 per km for the cost of underwater transmission cables. Two underwater cable landing points are located in Tofino and Ucluelet (×) and power grid connection point is located in Ucluelet (●). Each of the WEC facilities is composed of 28 Pelamis devices. The price of electricity is set at 20 cents per kW.

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. We will continue to update this section as we learn about new data sources and methods.

a. 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.

b. 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.

c. 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/

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.

Coastal Vulnerability Model

Summary

Faced with a changing climate and a growing intensity of human activities, it is imperative for coastal communities to understand how development and changes in the biological and physical environment can affect their exposure to storm-induced erosion and inundation. The InVEST Coastal Vulnerability model produces a qualitative estimate of such exposure, as well as a map of the location and size of human settlements. The Coastal Vulnerability Model maps the location of populations living near coastlines and their exposure to erosion and inundation during storms. The model computes an Exposure Index, which differentiates areas with relatively high or low exposure to erosion and inundation.

The model does not take into account coastal processes that are unique to a region, nor does it predict change in the shoreline. Model inputs, which serve as proxies for various complex shoreline processes that influence exposure to erosion and inundation, include: maps of coastal geomorphology, location of natural habitats (e.g., seagrass, kelps, wetlands, etc.), rates of net sea-level change, wind climate, and a depth profile that can be used as an indicator for surge level (default profile is the edge of the continental shelf). Outputs of the model are maps of coastal human populations and a coastal exposure to erosion and inundation index map. Outputs can be used to understand the relative contributions of different variables to coastal exposure and to highlight the protective services offered by natural habitats. This information can help coastal managers, planners, landowners, and other stakeholder identify regions of greater risk to coastal hazards, which can in turn inform development strategies and permitting. This is a "Tier 0" model.

Introduction

Coastal regions are constantly subject to the action of ocean waves and storms and naturally experience erosion and inundation over various temporal and spatial scales. However, coastal erosion and inundation can also threaten human populations, activities, and infrastructure, especially within the context of a changing climate and increases in the number of people living in these areas. Moreover, these increases in human pressure can lead to 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 bio-geophysical factors in increasing or decreasing coastal erosion and inundation in order to better plan for future development. It is also important to know how natural habitats can mitigate the forces responsible for coastal erosion and inundation so that any management action preserves the protection 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 estimate the relative vulnerability of coastal areas to erosion and inundation based on the bio-geophysical and social characteristics of a region. It is our aim to fill that gap with the Coastal Vulnerability model.

The Coastal Vulnerability model produces a map of the location and size of human settlements as well as a qualitative index of coastal exposure to erosion and inundation. The model does not value directly any ecosystem service, but ranks sites as having a relatively low, moderate or high risk of erosion and inundation, and maps the size of human population living near those sites. 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 they are formatted in a way that enables users to manipulate them as they see fit. They can be used to identify and distinguish segments of the coastline that are both populated by humans and vulnerable to natural hazard events. They can also help users become familiar with the characteristics of their coastal region, and help highlight the relative role of natural habitats at reducing exposure. Finally, outputs can be used to evaluate, in a simple way, how some management actions can increase or reduce exposure of human populations to the coastal hazards of erosion and inundation.

The Model

The InVEST Coastal Vulnerability model produces a Population map and an Exposure Index map. The population map shows human population density near the coastal region of interest. The Exposure Index map ranks the relative exposure of coastal regions and communities to erosion and inundation caused by large storms. These maps are constructed using a global population density map and five bio-geophysical variables that represent the natural biological and geomorphic characteristics of a region, the amount of expected net sea-level rise, and the relative effects of storms. Model outputs are relevant when computed for a relatively large coastal region. They can be used to locate populated areas that are more (or less) exposed to erosion and inundation than others.

How it works

The model creates the population and exposure index maps using a mixture of raster GIS and user input datasets of population and five bio-geophysical variables:

1. Geomorphology
2. Natural habitats (biotic and abiotic)
3. Net sea level change
4. Wind speed
5. Edge of continental shelf depth contour (or other depth contour that be used to estimate surge potential)

The outputs are mapped on the shoreline of the coastal region of interest with a user-defined spatial resolution (that can be as small as 250 meters). We provide a default global, high resolution shoreline database (Wessel and Smith 1996) to represent coastlines of most regions of the world. Below are details on how the output maps are created.

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 hazard events. Our Coastal Vulnerability model includes globally available population estimates for the coastal zone, which are derived from country-level census data. To obtain a raster dataset that shows the estimated number of people residing in each coastal grid cell, we overlay this input dataset with the shoreline data. We then assign each shoreline segment a population count by extracting the population count from the grid cell with the most overlap with the shoreline segment.

Exposure Index
We compute the physical exposure index by combining the ranks of the five 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 1). A rank of 3 represents average exposure. We built this ranking system based on methods proposed by Gornitz et al. (1990) and Hammar-Klose and Thieler (2001).

Table 1: List of Bio-Geophysical Variables and Ranking System for Coastal Exposure.

Rank	Very Low	Low	Moderate	High	Very High
Variable	1	2	3	4	5
Geomorphology	Rocky; high cliffs; fiord; fiard	Medium cliff; indented coast	Low cliff; glacial drift; alluvial plain	Cobble beach; estuary; lagoon; bluff	Barrier beach; sand beach; mud flat; delta
Natural Habitats	Coral reef; mangrove; coastal forest	High dune; marsh	Low dune	Seagrass; kelp	No habitat
Sea Level Change	Net decrease		±1		Net rise
Wind-wave Exposure	<10th Percentile	<25th Percentile	Average value	>75th Percentile	>90th Percentile
Surge Potential	No exposure	<25th Percentile	Average value	>75th Percentile	>90th Percentile

The model calculates the Exposure Index *E* for each shoreline segment as (Gornitz et al., 1990):

$$E = \sqrt{\frac{R_{BioPhysical} R_{SLR} R_{WaveExposure} R_{Surge}}{Count(R_{BioPhysical}, R_{SLR}, R_{WaveExposure}, R_{Surge})}} \quad \text{(Eq. 1)}$$

where:

- $R_{BioPhysical} = 0.75 R_{Geomorphology} + 0.25 R_{Habitats}$ is a combination of the exposure ranking of the geophysical and natural habitat variables. If $R_{Geomorphology}$ is lower than 2, then $R_{BioPhysical} = R_{Geomorphology}$. In other words, if the shoreline cannot be eroded after a single event, the presence/absence of natural habitats is not taken into account.
- $Count(R_{BioPhysical}, R_{SLR}, R_{WaveExposure}, R_{Surge})$ represents the sum of the variables that are taken into account. If any of the four (biophysical, SLR, wave exposure and surge potential) is not taken into account or is equal to zero, then it is removed from the count.

If, for example, SLR is not taken into account at a site of interest because it has the same value everywhere, then $R_{SLR} = 1$, and $Count = 3$ (assuming that habitats and forcing are still taken into account).

If there is no geomorphology layer, the Exposure Index is not computed: $E = 0$. The model will still compute $R_{Habitats}$, R_{SLR}, $R_{WaveExposure}$ and R_{Surge}, but it will not compute a final exposure index because geomorphology is assumed to be the most important variable that determines the potential for erosion/inundation at a site. Below, we present a more detailed description of the variables presented in Table 1.

Geomorphology

Rocky cliffs are less prone to erosion and inundation than bluffs, beaches, or deltas. Consequently, we adopted and hard-coded a relative ranking of exposure scheme based on geomorphology similar to the one proposed by Hammar-Klose and Thieler (2001). We provide in Appendix A a definition of the terms used in this classification, which applies mostly to the North American continent. We will expand this classification to more regions of the world in later versions of this model.

If the your geomorphology raster file has more categories than the ones presented in Table 1, we leave it to your discretion to reclassify your data to match our ranking system, as explained in the Data Needs section, and in Appendix B. We suggest however, that a high cliff is one that is above the 50-year surge level in a region, a low cliff is one that is at or below the 25-year surge level.

Natural Habitats

Natural habitats (marshes, seagrass beds, mangroves, coastal dunes) play a vital role in mitigating the effects of coastal hazards and decreasing the exposure of a coastal area and community. For example, large waves break on coral reefs before reaching the shoreline, mangroves and coastal forests dramatically reduce wave height 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. On the other hand, beaches with little to no biological habitats or sand dunes offer little protection to erosion and inundation. We developed the ranking proposed in Table 1 based on the fact that fixed and stiff structures that penetrate the water column (e.g., coral reefs, mangroves) and sand dunes are the most able to protect coastal communities. Flexible and seasonal structures, such as seagrass, reduce flows when they can withstand their force, and encourage accretion of sediments. Once again, we leave it to the users' discretion to separate sand dunes into high and low categories. We suggest however, based on Short and Hesp (1982), that 15m is an appropriate cut-off value to separate high (>15m) and low (<15m) dunes.

To compute a Natural Habitat exposure rank for a given shoreline segment, the model estimates whether a certain class of natural habitat (Table 1) is within a user-defined search radius from the segment. If more than one class of natural habitat are present and within the search radius from the segment, their ranks are allocated back to that segment. The final exposure rank of the segment is equal to the highest rank value of the various classes fronting the segment. Currently, the model only detects the presence of seagrass and kelp (see Section 2 and Appendix B for a description of how the model does this). We will include all the other habitat classes in a future model release.

Net Sea-Level Change
The relative net sea level rise/decrease along the coastline of a given region is the sum of global sea-level rise, local sea level rise (eustatic rise) and local land motion (isostatic rise). As indicated by Gornitz (1990), relative rise values between ±1 do not change current erosion or inundation trends, as they can be considered to be within modeling and measurement error range. In contrast, values smaller than -1 decrease the exposure, while values above +1 increase the exposure. Please consult Appendix B for suggestions of how to create this input.

Wind-wave exposure
We estimate the exposure of an area to wind-waves by computing the average wave power that winds can generate. For a given coastline segment, the model estimates the fetch distance over 16 equiangular sectors, with an accuracy of 1km, using a program developed by Finlayson (2005); maximum fetch distance is 50km and fetch distance overland is assumed to be zero. We combine this fetch distance with a user-input vector of the highest 5%wind speeds and corresponding direction. This vector should be derived from a record of wind speed and direction that is at least five years long (we will perform this extraction for users in the next model release). Appendix B provides some guidance on where to get this data. From fetch distance in each equiangular sector and top 5% wind speed and associated direction, wave height and period are computed as:

$$\begin{cases} H = \dfrac{0.24U^2}{g}\left[\tanh\left(4.14x10^{-4}\left(\dfrac{gF}{U^2}\right)^{0.79}\right)\right]^{0.572} \\ T = \dfrac{7.69U}{g}\left[\tanh\left(2.77x10^{-7}\left(\dfrac{gF}{U^2}\right)^{1.45}\right)\right]^{0.187} \end{cases} \quad \text{(Eq. 2)}$$

where
- F [m] is the average fetch distance in a particular sector
- U [m/s] is the maximum wind speed associated with a direction that falls in that sector (we will use all wind speeda in the next model release)
- g [m/s^2] is the acceleration of gravity

This expression does not differentiate between duration and fetch-limited conditions (USACE 2002; Part II Chap 2), and might under- or over-estimate wind-generated waves characteristics at a site. However, we believe that this simplification will not misrepresent the overall relative exposure of shoreline segments.

From estimated wave height and period, we compute average wave power by:

$$Power = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{16\pi}\rho g^2 H_n^2 T_n \quad \text{(Eq. 3)}$$

where:

- N is the number of equiangular sector for which the fetch is non-zero.
- H_n [m] is one of the 16 wave heights computed from a wind speed value in the sub-record
- T_n [s] is the wave period associated with H_n
- ρ is density of seawater, taken as 1,013 kg/m^3.

Final results are normalized by dividing each computed power value by the average of all wave power computed. Thus, a wave exposure value greater than one indicates a wave exposure greater than average (and a wave exposure <1 indicates less-than-average wave exposure).

<u>*Surge Potential*</u>

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 the region of interest, we estimate the relative exposure to storm surges by computing the length of the continental shelf fronting an area of interest. (For hurricanes, a better approximation might be made by considering the distance between the coastline and the 30 meters depth contour (Irish and Resio 2010)).

The tool that we use to perform this computation 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. Consequently, we offer the user the opportunity to define a maximum distance threshold over which shoreline segment within the area of interest will be deemed at low-risk of exposure to storm surge (see Data Needs section). We provide an example of how to estimate this distance in Appendix B.

<u>*Winds*</u>

Strong winds constitute a coastal hazard because they can generate high surges and/or high waves if they blow over an area for a long period of time. They can also move debris with tremendous force. In addition to an Exposure Index, we provide users with a map of shoreline segments that are exposed to or sheltered from strong winds. The classification is made by using a fetch filter: segments for which two or more of the 16 fetches do not exceed a user-defined distance are assumed to be sheltered. The relative exposure of segments is further refined by computing and mapping the Relative Exposure Index (REI; Keddy 1982). This index is computed, from the top 5% wind speed sub-record described in the Data Needs section, as:

$$REI = \sum_{n=1}^{16} U_n P_n F_n \qquad \text{(Eq. 4)}$$

where:

- U_n is the average wind speed, in meters per second, 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

Similarly to the wind-wave exposure, normalization of the REI is done by dividing the REI for a particular site by the average of all REI values.

Limitations and Simplifications

The primary technical limitation of the model is that we had to sacrifice accuracy for computational efficiency. Given the limitations of GIS, a raster-based calculation approach was taken, and any procedure that requires the model to evaluate the relative position of objects in space, such as the allocation of the presence of natural habitats to shoreline segments, has the potential to yield errors. Also, fetch distances are approximated to the nearest kilometer. Furthermore, while the global dataset of population that we are using, the GRUMP dataset (CIESIN and CIAT 2005), is advantageous because it is derived from census data rather than predictive models, it only represents an estimate of the resident population in all areas. The GRUMP data does not provide an estimate of the daily or seasonal distribution of population that often varies widely in coastal areas where tourism accounts for a large number of temporary residents. If better data are available for your region of interest, you can substitute these for the coarser global data.

Beyond technical limitations, the Exposure Index also has theoretical limitations. One of the main limitations is that we simplified the numerous natural characteristics and extremely complex coastal processes happening in a region into five variables and five exposure categories. For example, the model does not distinguish between sand and mixed sand beaches; nor does it take into account the slope of bluffs. More importantly, the model does not consider any hydrodynamic or sediment transport processes. Consequently, we assume that regions that belong to the same geomorphic exposure class behave in a similar way. Also, 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 1. For example, the relative exposure to waves and wind will have the same weight whether the site under consideration is a sand beaches or a rocky cliff.

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 fields and does not take into account any large-scale sediment transport pathways that may exist in a region of interest.

Data Needs

The model uses an interface to input all required and optional data, and as is outlined in this section, it outputs a population and an exposure index map. The population map is always produced, but the user has the option of uploading any or all of the variables in Table 1 to compute the Exposure Index map. This index can be created with fewer than the five variables required as input. However, the model always computes fetch distances based on the required input land polygon and polyline (inputs 2 and 3), and it outputs an index value of -1 if a geomorphology layer (input 15) is not provided. Here we outline the options presented to the user via the interface, and the content and format of the required and optional input data used by the model. More information on how to fill the input interface or how to obtain data is provided in Appendix B.

1. **Workspace Location (required).** The user is required to specify a workspace folder path. We recommended creating a new folder for each run of the model. For example, by creating a folder called “RunWholeCoast” within the “CoastalVulnerability” folder, the model will create “intermediate” and “output” folders within this “RunWholeCoast” 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\CoastalVulnerability\RunWholeCoast

2. **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 file is provided as default (Wessel and Smith, 1996), but other layers can be substituted.

   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

3. **Land Polyline (required)**. This input should have the same shape as the *Land Polygon* (input 2), and must have a feature geometry of polyline instead of polygon.

   Name: File can be named anything, but no spaces in the name
   File type: polyline shapefile (.shp)
   Sample path (default): \InVEST\Base_Data\Marine\Land\global_polyline.shp

4. **Area of Interest (AOI) (required).** The user must create a polygon feature layer that defines the Area of Interest (AOI). An AOI instructs the model where to clip the *Land Polygon* and *Land Polyline* input data (inputs #2-3) in order to define the spatial extent of the analysis.

   If the user is including the Surge Potential variable in the computation of the exposure index, the depth contour specified in input #13 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\CoastalVulnerability\Input\AOI_BarkClay.shp

5. **Cell Size (meters, required).** This input determines the spatial resolution at which the model runs and the resolution of the output maps. To run the model at the minimum 250 x

250 meters grid cell scale, the user should enter “250”. A larger grid cell will yield a lower resolution, but a faster computation time.

Name: A numeric text string (positive integer)
File type: text string (direct input to the ArcGIS interface)
Sample (default): 250

6. **Projection (required)**. The model will project or re-project most of the input data in order to keep the model’s various distance calculations consistent. This projection input must have meters as the units and use a WGS84 datum.

   File type: projection files provided by ArcGIS (.prj)
   Sample path: Coordinate Systems\Projected Coordinate Systems\UTM\WGS 1984\ WGS 1984 UTM Zone 10N.prj

7. **Indices Table (required).** The user must provide a summary table to instruct the model on various parameters necessary to calculate the exposure index. While numbers can be modified -- adding to, deleting, or rearranging the order of cells may produce erroneous results. The model expects values to be in these specific cells with the exception of the geomorphology listing (yellow) where the number of entries can be increased or decreased. More information on how to fill this table is provided in Appendix B (this chapter).

   Table Names: File can be named anything, but no spaces in the name
   File type: *.xls or .xlsx (if user has MS Excel 2007 or newer)
   Sample:
   InVEST\CoastalVulnerability\Input\ExposureIndexParameters_WCVI.xls\Indices$

1) NATURAL HABITAT	DISTANCE (meters)
kelp	1500
eelgrass	500
Minimum area threshold (m^2) to include habitat	75
3) SURGE POTENTIAL	DISTANCE (meters)
Maximum search distance from continental shelf to coastline within the user-specified AOI	145000
3) FETCH FILTER	DISTANCE (meters)
Threshold where if two or more of the 16 fetch directions do not exceed this distance, the model will assume that the shoreline segment is not exposed to the open ocean	10000
INSTRUCTIONS	
Note: Values in this table are parameters for the Coastal Protection model's exposure index. While numbers can be modified – adding to, deleting, or rearrange the order of cells may produce erroneous results. The model expects values to be in the cells outlined by this template, with the exception of the geomorphology listing where the number of entries can be increased or decreased.	

5) GEOMORPHOLOGY	ID	RANK
Placeholder	1	3
Rock Platform	2	2
Rock Cliff	3	1
Rock with Gravel Beach	4	2
Rock, Sand and Gravel Beach	5	3
Rock with Sand Beach	6	3
Gravel Beach	7	4
Gravel Flat	8	4
Sand and Gravel Beach	9	4
Sand Beach	10	5
Sand and Gravel Flat	11	5
Sand Flat	12	5
Mud Flat	13	5
Estuary, Marsh or Lagoon	14	4
Channel	15	4
Man-made	16	3

8. **Population Raster (required).** A raster layer is required 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 raster data 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

9. **Land Area Filter (kilometers squared, optional).** All landmasses within the AOI are included in fetch calculation, but this input instructs the model to filter out from the output calculation land masses (islands) with an area less than the value specified (in km^2). For example, if the user enters "5", the model will only produce outputs for landmasses that have an area greater or equal to $5km^2$. More information on how to fill this input cell is provided in Appendix B.

   This input should be left blank if (1) the user does not wish to filter out any land masses or (2) the user selects a land polygon and polyline (inputs #2-3) that is different from the default layers provided in the directory "\InVEST\Base_Data\Marine\Land\".

   Name: A numeric text string (positive integer)
   File type: text string (direct input to the ArcGIS interface)
   Sample (default): 5

10. **Natural Habitat: Kelp (optional)**. This input layer is used to compute a Natural Habitat ranking for each shoreline segment (see Table 1), and should consist of the location of kelp beds (which will be clipped by the model within the AOI, input #4). The distance at which this layer will have a protective influence on coastline can be modified in the indices tables (input #7).

    Name: File can be named anything, but no spaces in the name
    File type: polygon shapefile (.shp)
    Sample path: \InVEST\CoastalVulnerability\Input\kelp_CRIMS.shp

11. **Natural Habitat: Seagrass (optional)**. This input layer is used to compute a Natural Habitat ranking for each shoreline segment (see Table 1), and should consist of the location of seagrass beds (which again will be clipped by the model within the AOI, input #4). The distance at which this layer will have a protective influence on coastline can be modified in the indices table (input #7).

    Name: File can be named anything, but no spaces in the name
    File type: polygon shapefile (.shp)
    Sample path: \InVEST\CoastalVulnerability\Input\seagrass_CRIMS.shp

12. **Sea Level Rise: Polygon Indicating Net Rise or Decrease (optional)**. 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 "RANK" that orders the net change values according to

Table 1. More information on how to create this polygon is provided in the Marine InVEST Frequently Asked Questions (FAQ) document at http://invest.ecoinformatics.org, and in Appendix B (this chapter).

Name: File can be named anything, but no spaces in the name
File type: polygon shapefile (.shp)
Sample path: \InVEST\CoastalVulnerability\Input\SeaLevRise_WCVI.shp

13. **Surge Potential: 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 #4).

    Names: File can be named anything, but no spaces in the name
    File type: polygon shapefile (.shp)
    Sample path: \InVEST\CoastalVulnerability\Input\continentalShelf.shp

14. **Wind-Wave Exposure: Wind Vector List (optional).** This input is used to compute the Wind-wave Exposure ranking of each shoreline segment (Table 1). It consists of a text file (.txt) with two lines for the speed (in meters/seconds, line 1) and associated direction (in degrees (0-360), line 2) of the highest 5% winds measured near the AOI (input #4). Input values should represent the 5 to 10 minutes averaged wind speeds; they should not correspond to gust speeds.

    Table Names: File can be named anything, but no spaces in the name
    File type: *.txt
    Sample: \InVEST\CoastalVulnerability\Input\WindVectorList.txt

15. **Geomorphology: Shoreline Type (required)**. This input, of geometry type "polyline", is used to compute the Geomorphology ranking of each shoreline segment (Table 1). It does not have to match the land polyline input (input #3), but must resemble it as closely as possible. Additionally, the polyline shapefile must have a field called "ID" that identifies the various shoreline types with a number. The user must assign a corresponding rank value to each ID in the indices table (input #7). 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\CoastalVulnerability\Input\Shorezone_VI.shp

Running the Model

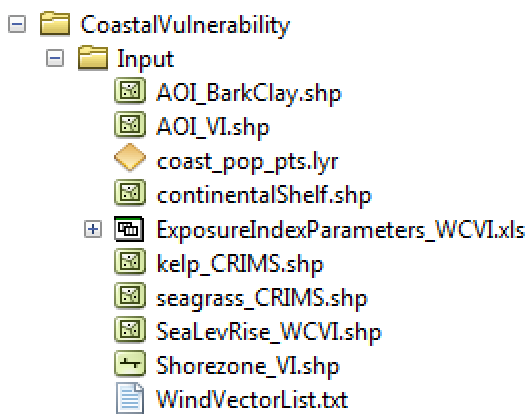
Setting up workspace and input folders

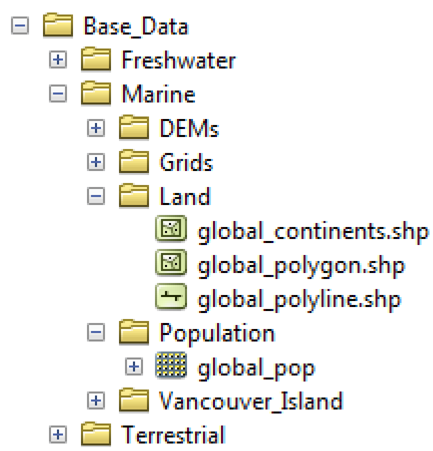
These folders will hold all input, intermediate and output data for the model. As with all folders for ArcGIS, these folder names must not contain any spaces or symbols. See the sample data for an example.

Note about terminology used here: 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\CoastalVulnerability* folder holds the main working folder for the model and all other associated folders. Within the *CoastalVulnerability* 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.

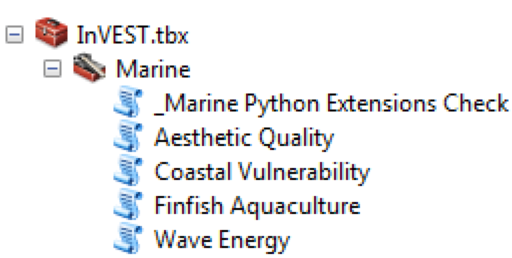
The following image shows the sample inputs (on the left) and base data (on the right) folder structures and accompanying GIS data. We recommend using this folder structure as a guide to organize your workspaces and data. Refer to the screenshots below for examples of folder structure and data organization.

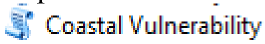




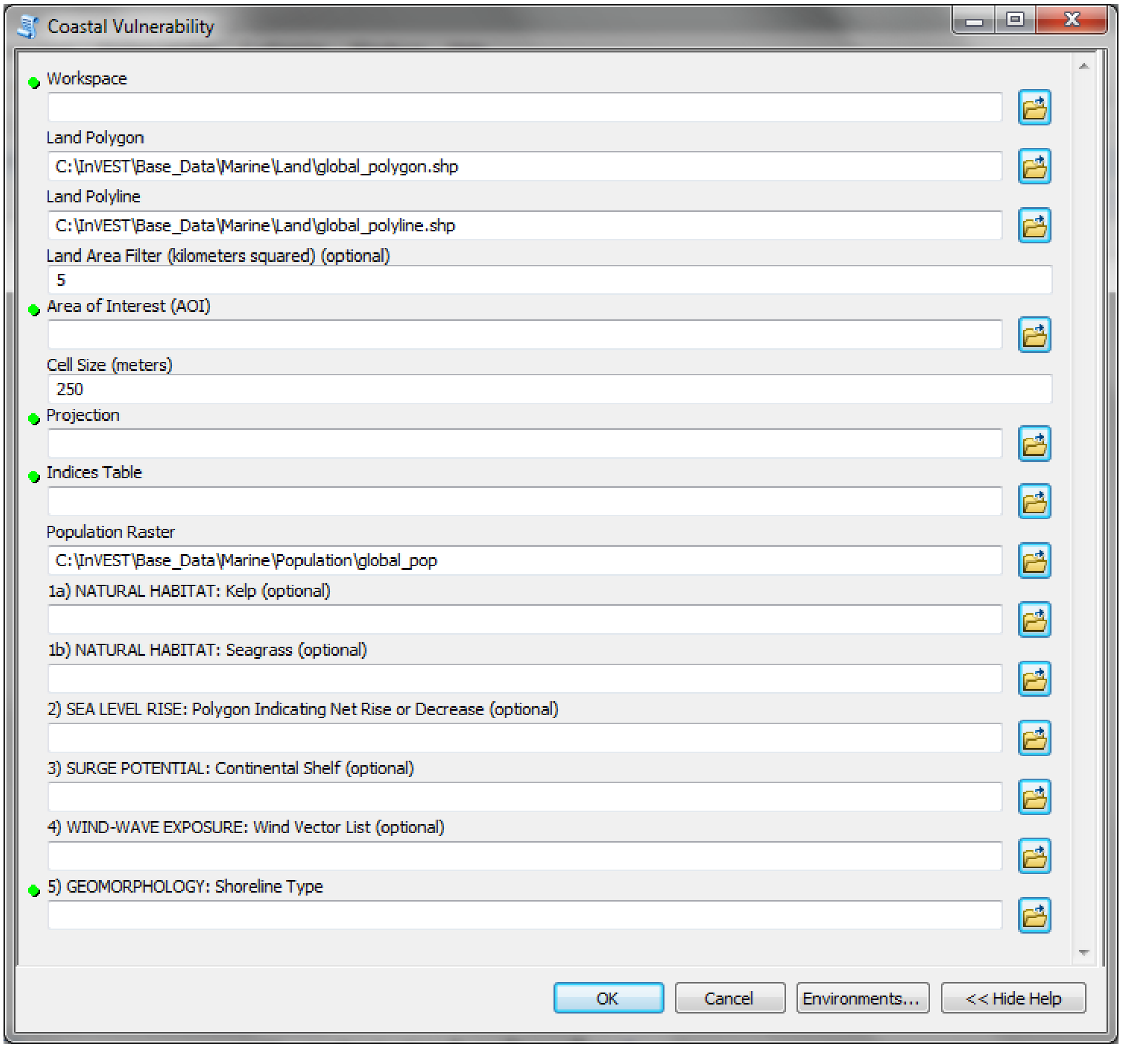
Creating a run of the model
The following example of setting up the Coastal Vulnerability model uses the sample data provided with the InVEST download. The instructions and screenshots refer to the sample data and folder structure supplied with the InVEST installation package. It is expected that users will have location-specific data to use in place of the sample data. These instructions provide only a guideline on how to specify to 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.

a. Click the plus symbol  next to the InVEST toolbox.

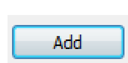


b. Expand the Marine toolset and click on the Coastal Vulnerability script 

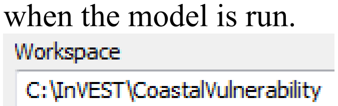
to open the model.



c. Specify the Workspace. Click on the Open Folder button  and path to the *\InVEST\CoastalVulnerability* folder. If you created your own workspace folder (Step 1b), then select it here.

Click on the *CoastalVulnerability* folder and click 

to set the main model workspace. This is the folder in which you will find the intermediate and final outputs when the model is run.



d. Specify the Land Polygon. The model requires a land polygon shapefile to define the shoreline for the analysis. This shapefile will be supplied in the model window for you.

Land Polygon

C:\InVEST\Base_Data\Marine\Land\global_polygon.shp

e. Specify the Land Polyline. The model requires a land polyline shapefile to define the shoreline for the analysis. This shapefile will be supplied in the model window for you.

Land Polyline

C:\InVEST\Base_Data\Marine\Land\global_polyline.shp

f. Specify the Land Area Filter (Optional). If you select this option, the model requires a land area filter parameter. The default value is given as 5 square kilometers. You can change this value by directly typing into the text box and entering another value.

Land Area Filter (kilometers squared) (optional)

5

g. 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 *AOI_BarkClay.shp* shapefile supplied in the sample data. You can create an AOI shapefile by following the *Creating an AOI* instructions in the **Getting Started** section. Open the *\InVEST\CoastalVulnerability\Input* data folder. Select the AOI_BarkClay.shp shapefile and click Add to make the selection.

Area of Interest (AOI)

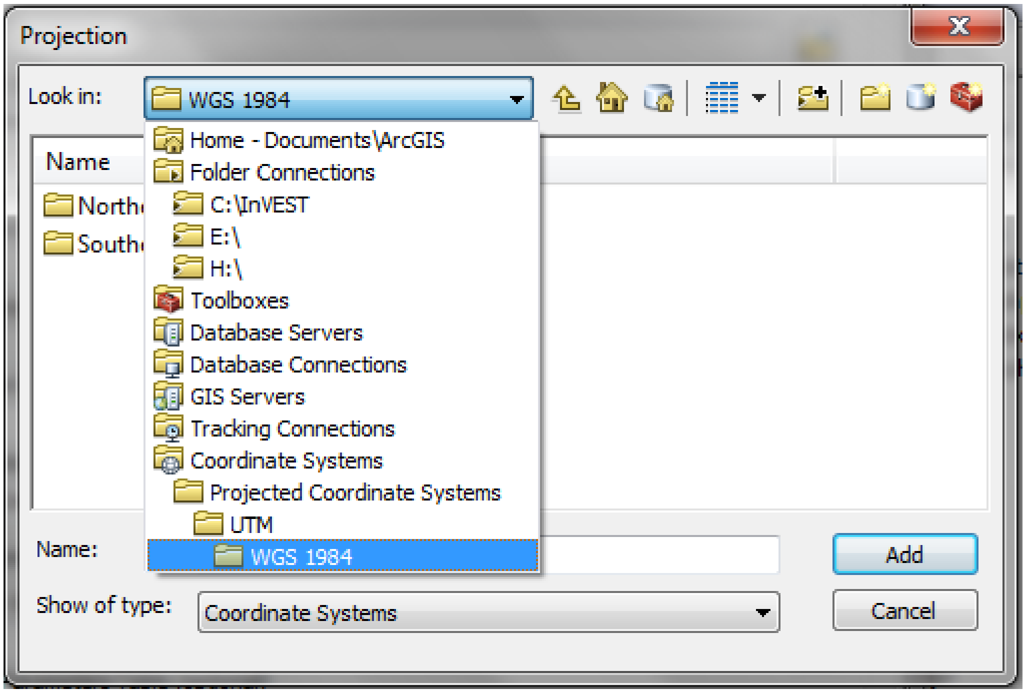
C:\InVEST\CoastalVulnerability\Input\AOI_BarkClay.shp

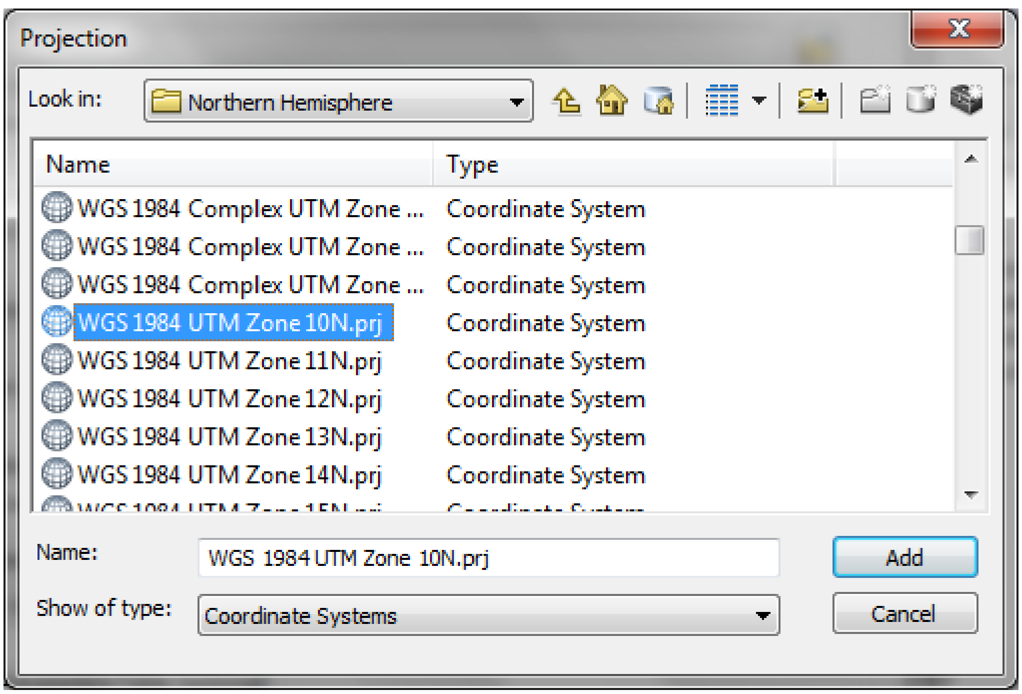
h. Specify the Cell Size. The model requires a cell size for the raster analysis. The default cell size is 250 meters. You may change this value by entering a new value directly into the text box.

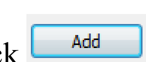
Cell Size (meters)

250

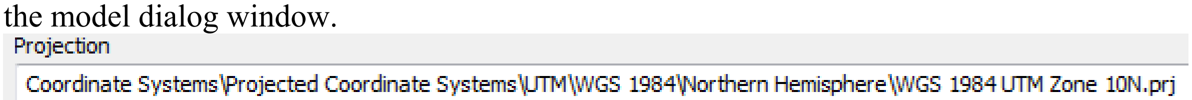
i. Specify the Projection. The Projection file is specified to set the projection and coordinate information necessary to run the overlay analysis. Open the Coordinate Systems folder near the bottom of the Look In list and path to the *UTM\WGS 1984* folder.





Select the WGS 1984 UTM Zone 10N.prj projection file and click 

to add it to the model dialog window.



It is assumed that all of your input data are in the same projection and coordinate systems with matching datum. If you need to (re)project your data see the Projection section in the **FAQ document** (http://invest.ecoinformatics.org) and/or the **Getting Started section**.

j. Specify the Indices Table. The model requires a table of exposure indices stored in a Worksheet in an Excel workbook file (.xls). See the **Data Needs** section for more information on creating and formatting these data. This worksheet will be supplied for you.

Click and path to the *\InVEST\CoastalVulnerability\Input* data folder. Double left-click on the Excel file Exp*osureIndexParameters_WCVI.xls* and select the worksheet *Indices$*.

Click Add to make the selection.

Indices Table

C:\InVEST\CoastalVulnerability\Input\ExposureIndexParameters_WCVI.xls\Indices$

Note: ArcMap and the model will not recognize the Excel sheet as valid data if it is added to the Data View. It is best to specify Excel data directly in the model dialog window using the Open folder and Add buttons and navigating to the data.

k. Specify the Global Population Raster. This is a global population raster with population assigned to each cell value. This raster will be supplied in the model window for you. Click and path to the *\InVEST\Base_Data\Marine\Population* folder. Select the *global_pop* raster and click Add to make the selection.

Population Raster

C:\InVEST\Base_Data\Marine\Population\global_pop

l. Specify the Natural Habitat: Kelp (Optional). The model can use an optional polygon shapefile that represents kelp habitat. Click and path to the *\InVEST\CoastalVulnerability\Input* data folder. Select the *kelp_CRIMS.shp* shapefile and click Add to make the selection.

1a) NATURAL HABITAT: Kelp (optional)

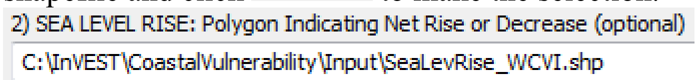
C:\InVEST\CoastalVulnerability\Input\kelp_CRIMS.shp

m. Specify the Natural Habitat: Seagrass (Optional). The model can use an optional polygon shapefile that represents seagrass habitat. Click and path to the *\InVEST\CoastalVulnerability\Input* data folder. Select the *seagrass_CRIMS.shp* shapefile and click Add to make the selection.

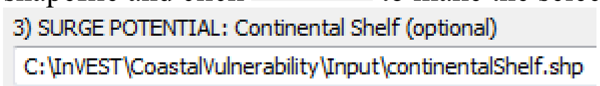
1b) NATURAL HABITAT: Seagrass (optional)

C:\InVEST\CoastalVulnerability\Input\seagrass_CRIMS.shp

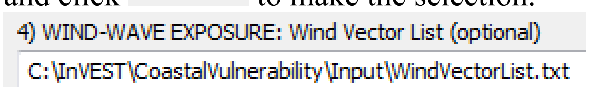
n. Specify the Sea Level Rise polygon (Optional). The model can use an optional polygon shapefile that represents sea level rise potential. Click and path to the *\InVEST\CoastalVulnerability\Input* data folder. Select the *SeaLevRise_WCVI.shp* shapefile and click to make the selection.



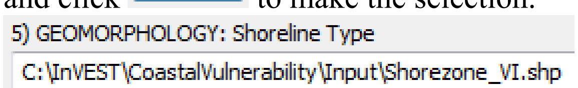
o. Specify the Surge Potential data (Optional). To represent surge potential, the model uses a continental shelf polygon shapefile. . Click and path to the *\InVEST\CoastalVulnerability\Input* data folder. Select the *continentalShelf.shp* shapefile and click to make the selection.



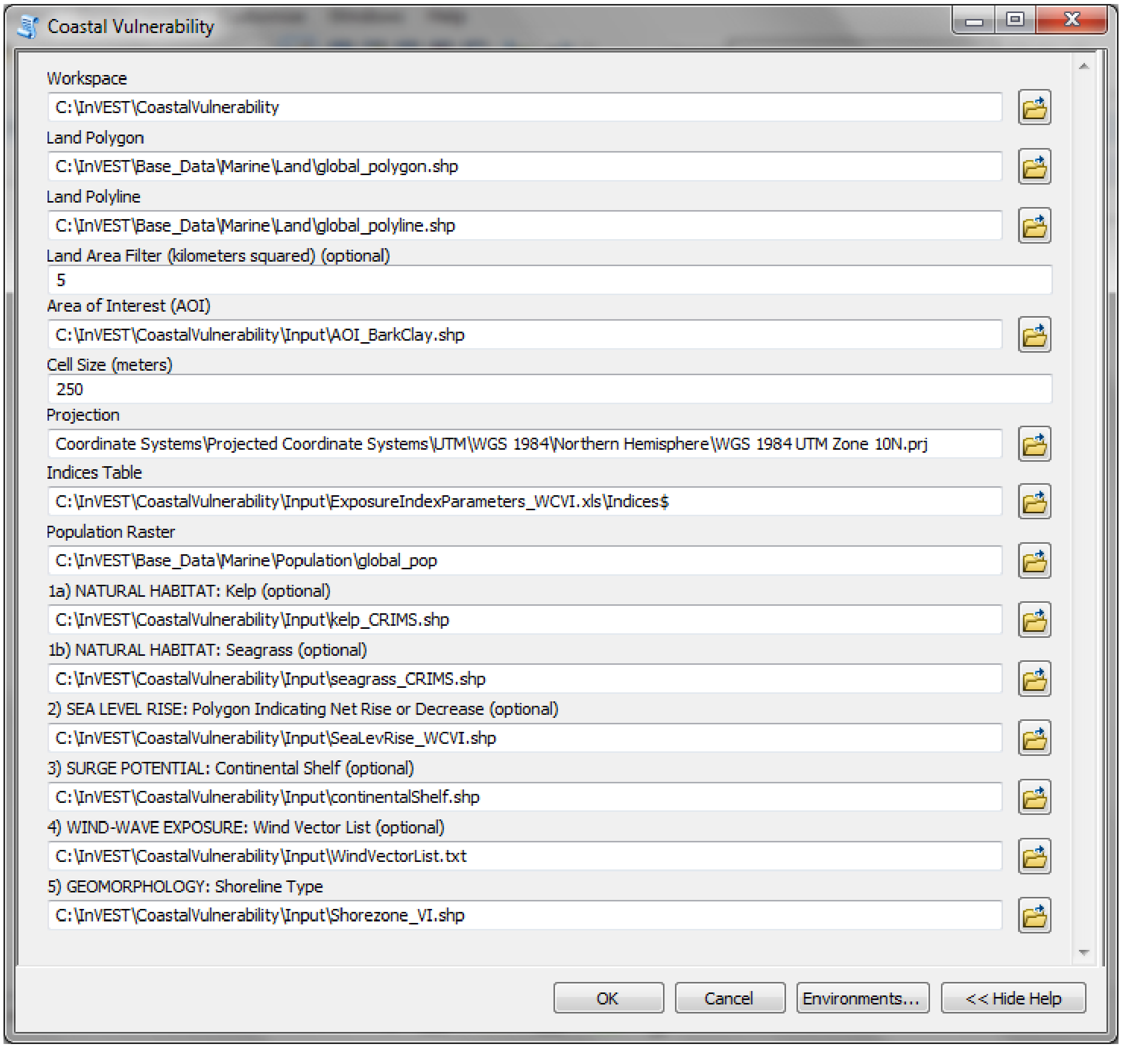
p. Specify the Wind-Wave Exposure data (Optional). The model can use an optional text file that represents wind exposure. See the **Data Needs** section for details on preparing these data and formatting the text file. Click and path to the *\InVEST\CoastalVulnerability\Input* data folder. Select the *WindVectorList.txt* textfile and click to make the selection.

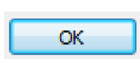


q. Specify the Geomorphology (Required). The model requires a polyline shapefile that represents shoreline geomorphology. Click and path to the *\InVEST\CoastalVulnerability\Input* data folder. Select the *Shorezone_VI.shp* shapefile and click to make the selection.

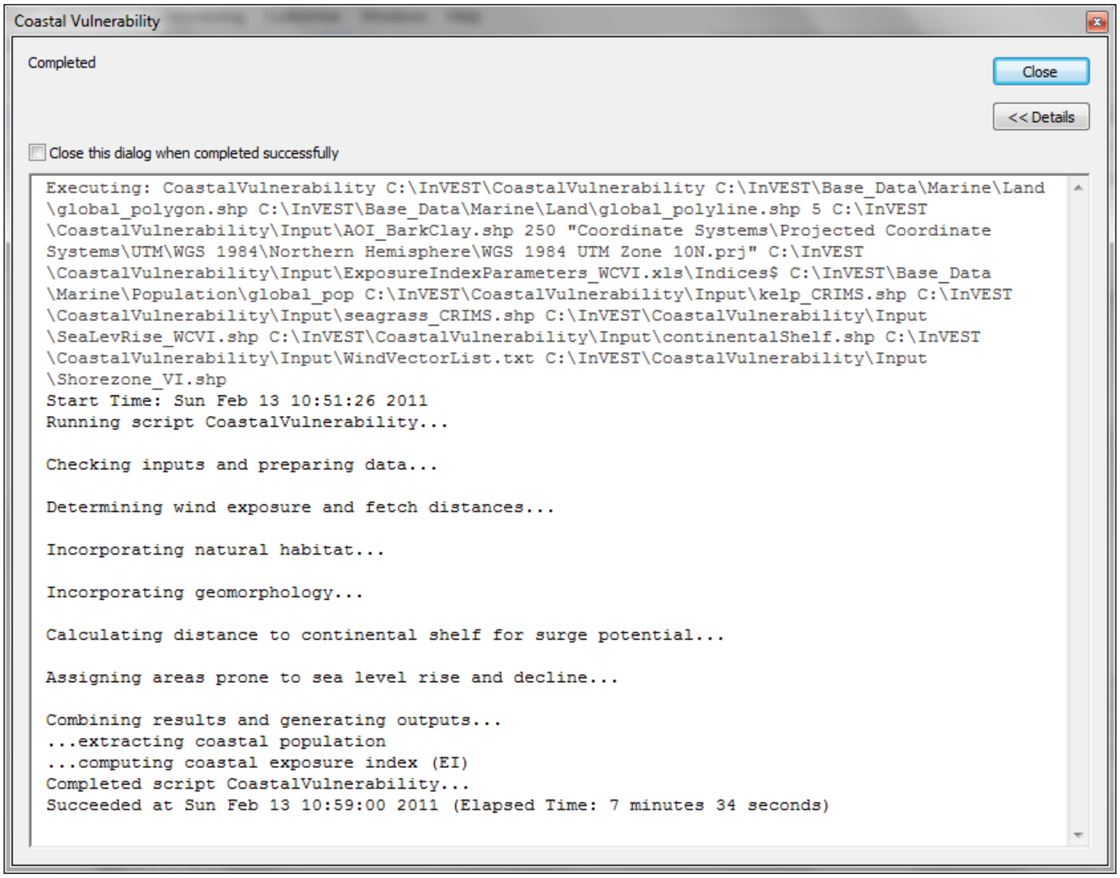


r. At this point the model dialog box is completed for a complete run (with all optional data for full exposure analysis) of the Coastal Vulnerability model.



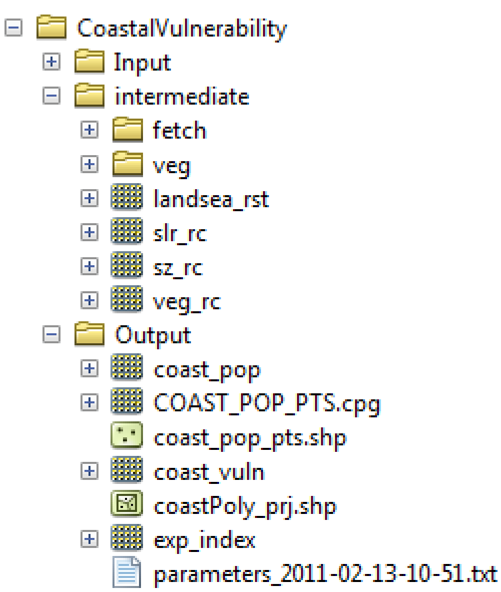
Click 

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.

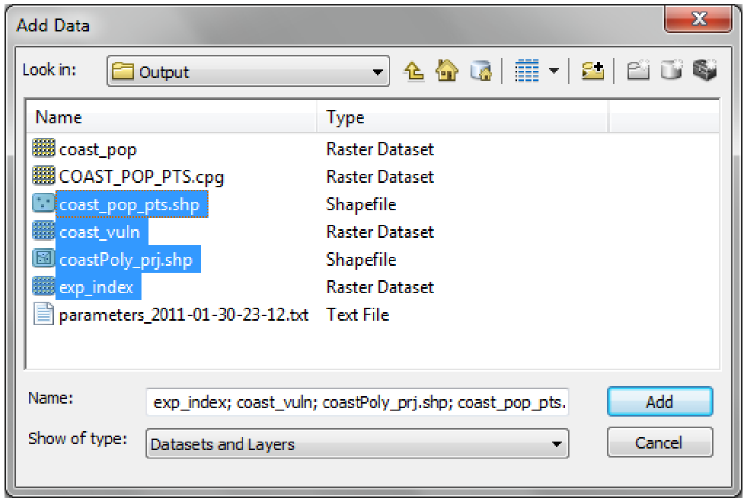


Viewing output from the model

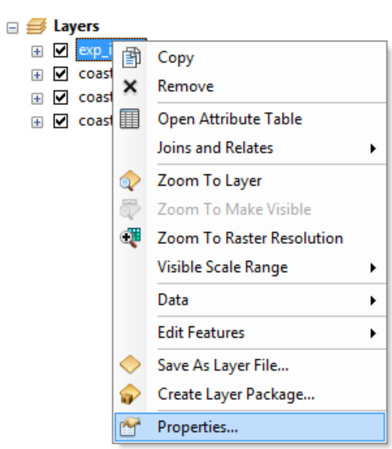
Upon successful completion of the model, two new folders called "*intermediate*" and "*Output*" will be created in the workspace. The *Output* folder, in particular, may contain several types of spatial data, each of which are described the Interpreting Results section.

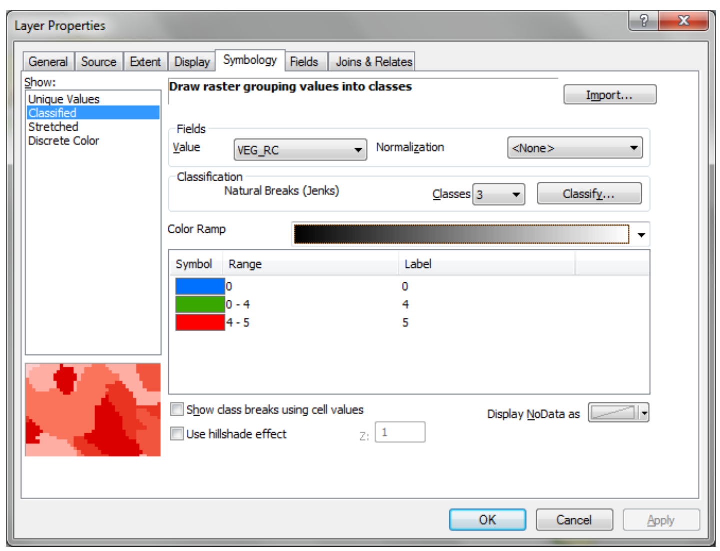


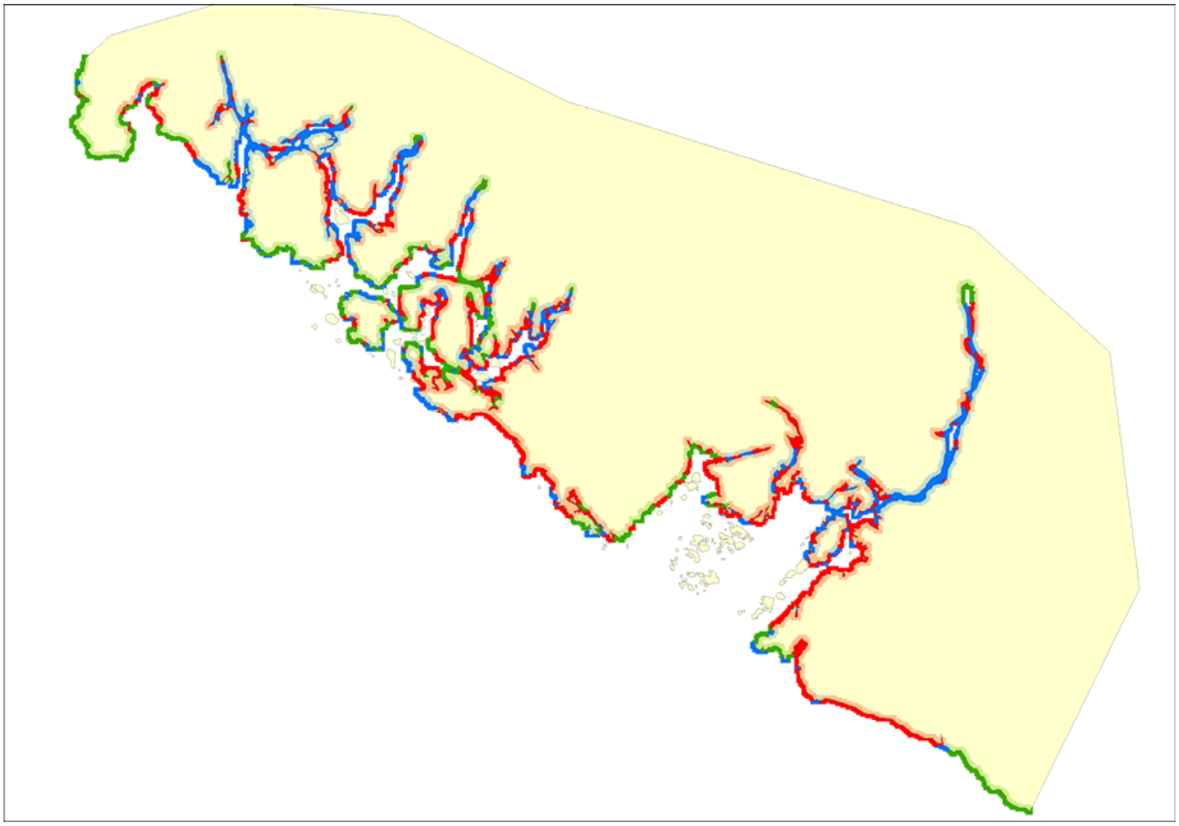
To view the output spatial data in ArcMap (from either the Intermediate or Output folders) click the Add Data button. , and select the four files highlighted in the figure below.



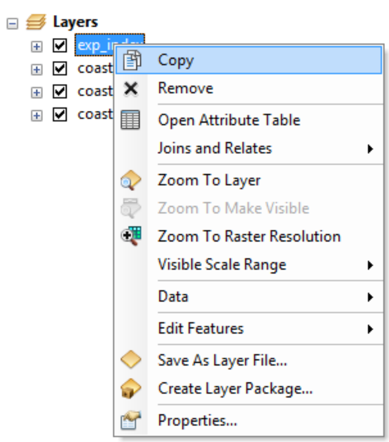
To navigate between the different fields contained in the "exp_index" outputs, or to change the symbology of a layer, right-click on the layer name in the table of contents, select "Properties", and then "Symbology". There are many options here to change the way the data appear in the map.

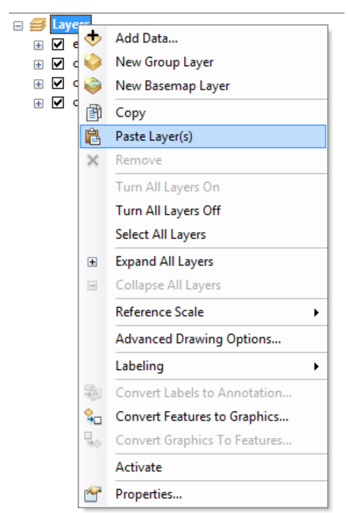




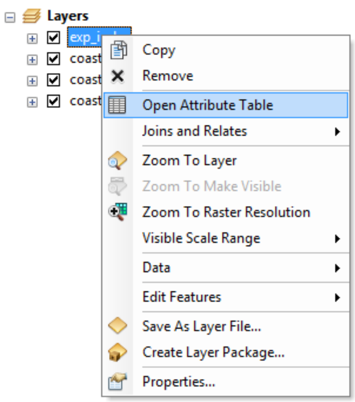


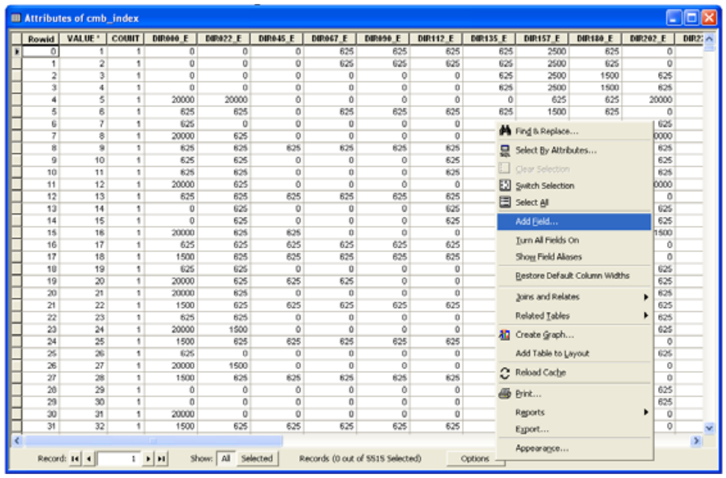
To navigate quickly between maps of output fields in the “exp_index” layer, we recommend you copy and paste the “exp_index” layer in the workspace:



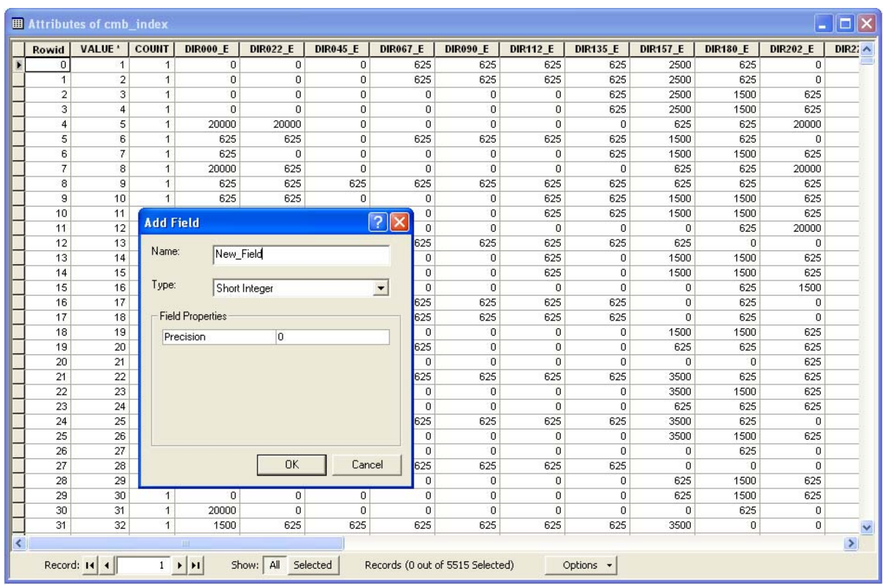


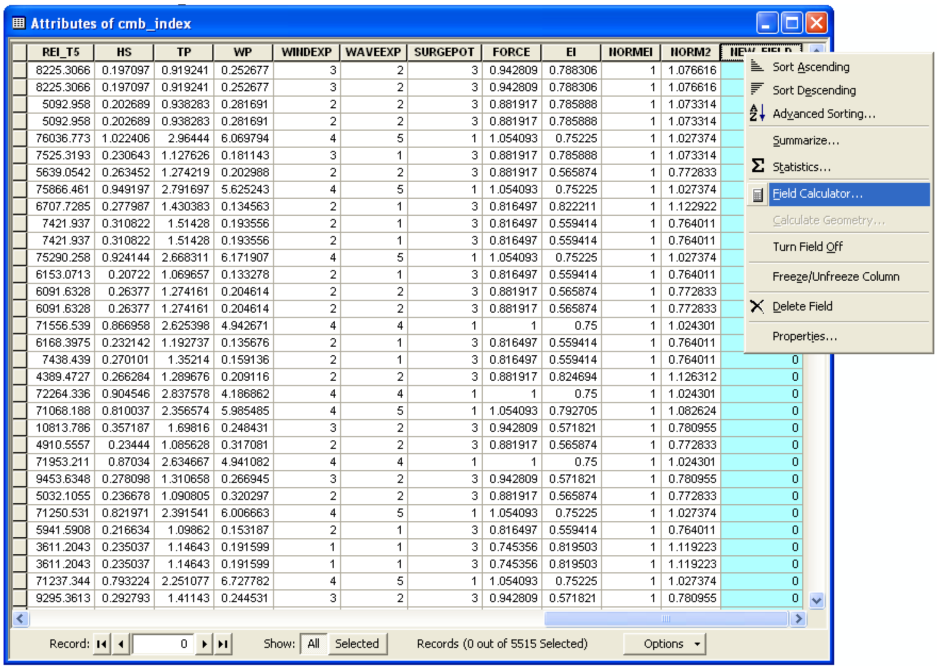
Finally, to generate a different map of outputs based on any other preferred relationship than the one presented in Equation (1) (see Gornitz 1990 for examples of other ways of computing the exposure index), we recommend creating a new field in the Attribute table:





Once the new field is created, it can be named "New_Index" (for example). After it is created, you can manipulate the various fields in any possible way using the field calculator:





We encourage the user to view as many fields in the outputs as necessary to develop an understanding of how the values of the different variables used to compute the exposure index change along the Area of Interest, and to view the optional outputs described in the Interpreting Results section.

Interpreting Results

Model Outputs

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

"Output" Folder

1. Output\ exp_index
   - This raster layer contains important statistics used to determine coastal exposure.
   - The raster contains a variety of fields, including:
     i. "EI" – the exposure index
     ii. "WINDEXP" – ranking (1-5) for wind exposure component of the index
     iii. "WAVEEXP" – ranking (1-5) for wave exposure component of the index
     iv. "SURGPOT" – ranking (0-5) for surge potential component of the index
     v. "BIOPHYS" – ranking (0-5) for biophysical component of the index
     vi. "GEOMORPH" – ranking (0-5) for geomorphology component of the index
     vii. "VEG_RC" – combine impact of all vegetation inputs
     viii. "SZ_RC" – the shoreline types allocated to each coastline segment
     ix. "COSTDIST_RC" – distance to continental shelf from each coastline segment
     x. "SLR_RC" – expected sea level rise rankings

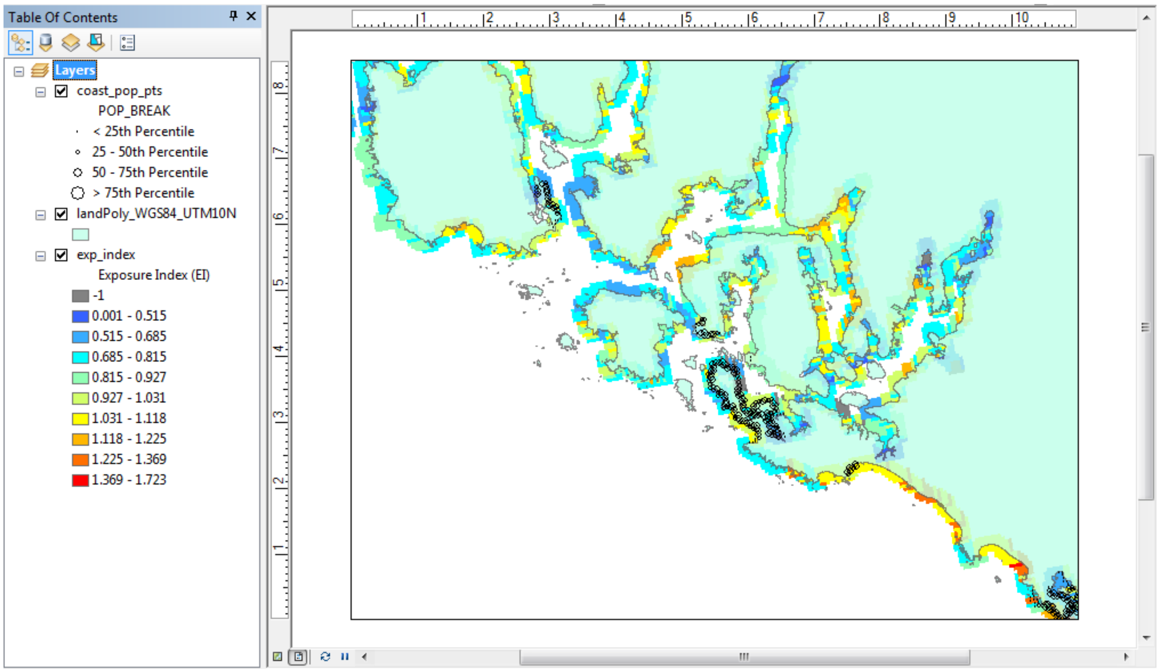
xi. "FFILT" – coastline segments with low (0) and high (1) exposure based on number of fetch directions exceeding a distance threshold
xii. "REI_T5" – relative exposure index as calculated from wind
xiii. "HS" – maximum wave height as calculated from wind
xiv. "TP" – wave period as calculated from wind associated with HS
xv. "WP" – wave power as calculated from wind

2. Output\ coast_vuln
   - This raster layer contains only values from the "EI" field of output #1 and is automatically symbolized when added to ArcMap.

3. Output\ coast_pop
   - This raster layer depicts population extracted from the global population input layer, but only for areas along the coast and within the user-specified area of interest.
   - The values this dataset represents are the number of people within each grid cell. The size of the grid cells is determined by the user.

4. Output\ coast_pop_pts.shp
   - The point feature layer contains points along the coastline only where people live.
   - This layer can easily be symbolized by importing the symbology from the file "\InVEST\CoastalVulnerability\Input\coast_pop_pts.lyr"

5. Output\ coastPoly_prj.shp
   - This polygon feature layer displays the clipped landmass within the AOI and is projected based on the projection specified by the user.
   - This layer is most useful when added to ArcMap and moved below all other output layers in the ordering hierarchy.



"intermediate" Folder

1. intermediate\fetch
   - This is a folder containing all 16 directional fetch calculations in raster format. Each direction has also been reclassed and expanded in order for the model to extract a fetch value for each coastline pixel.

2. intermediate\veg
   - This is a folder containing calculations for determining biogenic habitat's reach in terms of coastal protection. The reach distance of the two types of vegetation is set in the indices table (input #7). There will either be 1 or 2 files within this folder, depending on whether the user specifies both kelp and seagrass as inputs.

3. intermediate\veg_rc
   - This raster layer compiles the combine impact of all vegetation inputs. A value of 4 indicates areas where vegetation may impact coastal protection and a value = 5 means no impact.

4. intermediate\landsea_rst
   - This raster layer indicates areas of land (value = 1) and sea (value = 0) within the AOI and is used by the model to calculate fetch and distance to shelf.

5. intermediate\slr_rc
   - This raster layer indicates the expected sea level rise rankings within the AOI and is used as part of the exposure index.

6. intermediate\sz_rc

- This raster layer depicts how shoreline types were allocated to various sectors of coastline based on the polyline input #15 (geomorphology).

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.

Appendix A

Here we provide definitions for the terms presented in the geomorphic classification in Table 1. Some of these are from Gornitz et al. (1997) and USACE (2002).

<u>Alluvial Plain</u>: A plain bordering a river, formed by the deposition of material eroded from areas of higher elevation.

<u>Barrier Beach</u>: 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.

<u>Beach</u>: 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.

<u>Bluff</u>: A high, steep back or cliff

<u>Cliffed Coasts</u>: 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.

<u>Delta</u>: 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.

<u>Estuary Coast</u>: 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.

<u>Fiard</u>: Glacially eroded inlet located on low-lying rocky coasts (other terms used include sea inlets, fjardur, and firth).

<u>Fjord</u>: A narrow, deep, steep-walled inlet of the sea, usually formed by entrance of the sea into a deep glacial trough.

<u>Glacial Drift</u>: 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 strength.
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.

Appendix B

The model requires large-scale geo-physical, 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, we propose various sources for the different data layers that are required by the model, and we suggest methods to fill out the input interface discussed in the Data Needs section. We recommend that the user 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, we use population data from the Global Rural-Urban Mapping Project (GRUMP, http://sedac.ciesin.columbia.edu/gpw). This dataset contains global estimates of human populations in the year 2000 in 30 arc-second (1km) grid cells. You can use your own, more detailed and/or recent census data, and we encourage you to use recent fine-scale population maps, 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 you to substitute it with another layer.

To compute the Geomorphology ranking, the user 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 (www.geobc.gov.bc.ca). If such a database is not available, we recommend building such a database from site surveys information, aerial photos, geologic maps, or satellites images (using Google or Bing Maps, for example).

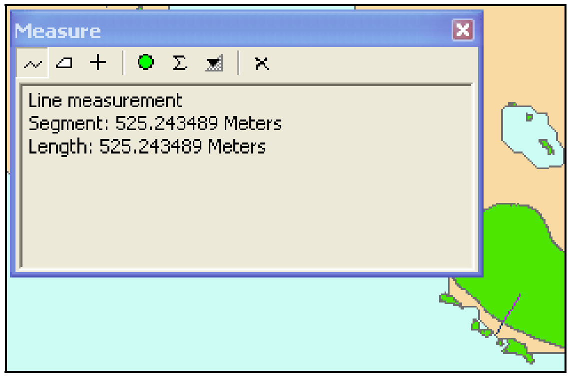
In addition to the geomorphology layer, the user must fill a Geomorphology table in the Indices Tables Excel file (*ExposureIndexParameters.xls*, see input 7 in the Data Needs Section). The table is used by the model to assign a geomorphology exposure ranking based on the different geomorphic classes identified. To fill out the Geomorphology table, we recommend pasting into the first two columns all the geomorphic classes that are in your AOI and their corresponding ID's in the geomorphology GIS layer. Then you need to assign a rank to those ID's, based on the classification we presented in Table 1. There is no limit to the number of unique geomorphology layer types, but the table must start with ID=1 and follow with consecutive integers. A placeholder may be used to preserve the sequence.

Habitat Data Layer

The natural habitat maps (inputs 10 and 11 in the Data Needs Section) should provide information about the location and types of coastal habitats described in Table 1. In some parts of the west coast of the United States and Canada, such a map can be built from a database called Shorezone (www.geobc.gov.bc.ca). If such a database is not available, we recommend building it from site surveys information, aerial photos, or even satellites images (using Google or Bing Maps, for example).

The model determines the presence or absence of various user-specified natural habitats in the AOI by estimating the fetch distance over 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 knows that the patch fronts that segment. To assign a natural habitat ranking to that segment that takes into account the beneficial effect of the presence of this habitat, we ask the user to input a maximum distance of influence into the Natural Habitat table in the Indices Tables Excel file (*ExposureIndexParameters*, see input 7 in the Data Needs Section). We assume that natural habitats that are fronting a segment but are further away from the segment that the user-defined distance 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, we recommend loading the various habitat layers (inputs 10 and 11) as well as the global polygon layer (input 1) and then zooming into the area of interest (AOI, input 4). Then, we recommend using the "distance tool" to 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 geospatial referencing of the input layer. The example below gives an example of such measurement when seagrass beds are considered (green patches).



Wind Data

To estimate the importance of wind-generated waves, and to express the relative exposure of shoreline segments to wind forces, we require the user to upload the highest 5% wind speeds and associated direction measured in the vicinity of the AOI. We recommend that this input data be derived from at least 10 years of measurement. The National Data Buoy Center (http://www.ndbc.noaa.gov/) provides links to maritime buoys that are located in, and maintained by, various countries around the world. Other sources of data are local airports, or Weather Underground (http://www.wunderground.com/), which provides records of measurements taken by certified, and sometimes uncertified, sources.

As described in "The Model" section "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 a user-input estimate of a fetch distance cutoff to use, based on the AOI under consideration. To provide that distance, we recommend using the "distance tool" on the global polygon layer (input 1), zoomed into the AOI, to determine that distance.

Sea Level Change

As mentioned earlier, the model requires a map of net rates of sea level rise or decrease in the AOI. Such information can be found in reports or publications on Sea Level Change or Sea Level Rise in the region of interest. Otherwise, we suggest that the user 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 gage measurements to be used in the context of sea level rise is the Permanent Service for Sea Level (http://www.psmsl.org/). 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, we suggest 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 "*Projected sea level changes for British Columbia in the 21st century*" by B. Bornhold (2008).

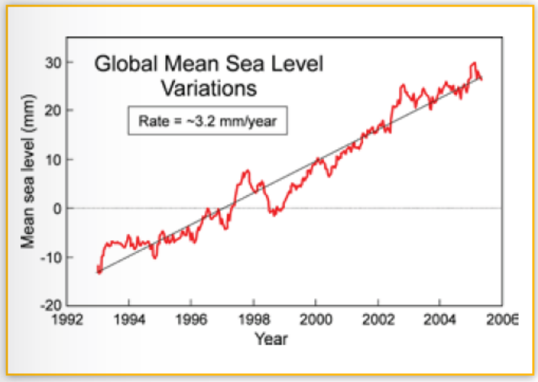


Figure 3: Global mean sea level change since 1993 based on satellite altimetry

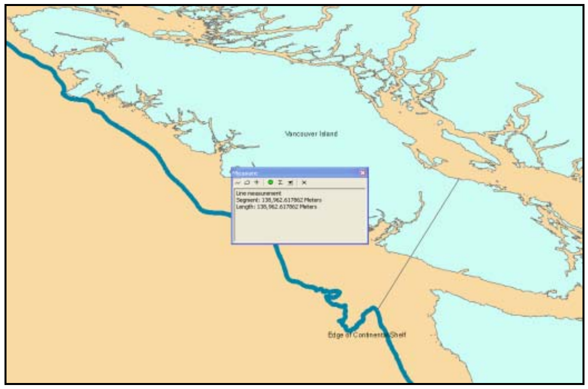
Create a Sea Level Change GIS Layer

Users can create their own polygon used as the sea level change input to the model. To create the feature class, 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. 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 for specific regions are created, the user must create an attribute field called "RANK" and populate it with either a value of 1, 3, or 5 indicating the net change values according to Table 1. For more information on how to create a Sea Level Change layer, see the Marine InVEST FAQ document at http://invest.ecoinformatics.org.

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. If the user feels that segments that are within the AOI are too far from the open ocean to be affected by surges, in the *ExposureIndexParameters* table we offer the possibility of limiting the search distance to areas that are closer than the maximum distance.

To fill the Surge Potential table in the *ExposureIndexParameters*, we recommend loading the global polygon layer (input 1) and the continental shelf (or other preferred depth contour, input 13), zoomed in to the AOI, and using the "distance tool" (see previous section). An example of such measurement is given below in the case of Vancouver Island, for which we believed that storm surges on the west coast would not affect the east coast of the island.



References

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

Finlayson, D. 2005, fetch program, USGS. Accessed February 2010, from http://sites.google.com/site/davidpfinlayson/Home/programming/fetch

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

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*".

MARINE FISH AQUACULTURE

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. This is a "Tier 1" model.

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, and will include the latter in the next model release (May 2011). Intermediate 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.

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.

1) 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:

$$W_{t,y,f} = \left(aW_{t-1,y,f}{}^{b} \cdot T_{t,f}\tau\right) + W_{t-1,y,f} \qquad \text{(Eq. 1)}$$

where a ($g^{1-b}day^{-1}$) and b (non-dimensional) are growth parameters, $T_{t,f}$ is daily water temperature (◦C) at farm f, and τ τττ(0.8◦C^{-1}) is a fixed scalar that represents the doubling of biochemical rates in fish when temperature increases by 8-9^{o}C. 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.

2) 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*:

$$TPW_{f,c} = W_{t_h,y,f} \cdot d \cdot n_f e^{-M*(t_h - t_o)} \qquad \text{(Eq. 2)}$$

where *Wthyf* is the weight at date of harvest t_{h},y on farm *f* from Eqn.1, *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*(th-t0)* is the daily natural mortality rate *M* experienced on the farm from the date of outplanting (t_0) to date of harvest (t_h).

3) Restocking

The previous 2 steps describe how fish growth is modeled 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.

4) 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:

$$NPV_{f,c} = TPW_{f,c}[p(1-C)] \cdot \frac{1}{(1+r)^t} \qquad \text{(Eq. 3)}$$

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%.

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), but future iterations will allow for monthly or yearly temperature inputs.

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

[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.

2. **Finfish Farm Location (required).** A GIS polygon or point dataset, with a latitude and longitude value and a numerical identifier for each farm. 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.

   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.).
   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".
   Sample data set: \InVEST\Aquaculture\Input\Finfish_Netpens.shp

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). The farm identifier name specified here must exactly match (including upper/lowercase and spaces) the name of one of the column headings from input #2. Additionally, the numbers underneath this farm identifier name must be unique integers for all the inputs (#2, 6, & 7).

   Names: A string of text that is 8 characters max
   File type: text string (direct input to the ArcGIS interface)
   Sample: FarmID

4-5. **Fish growth parameters (required, defaults provided)** Default a (0.038 g/day) and b (0.6667 dimensionless units) 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 (Eqn. 1) 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

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 (www.nodc.noaa.gov/SatelliteData/pathfinder4km/available.html) and Canada's Department of Fisheries and Oceans Oceanographic Database (http://www.mar.dfo-mpo.gc.ca/science/ocean/database/data_query.html). 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. For clarification, please refer to the sample temperature dataset in the InVEST package (Temp_Daily.xls).

Columns: The first two columns contain the number for that year (1-365) and day-month. 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.
Sample: \InVEST\Aquaculture\Input\Temp_Daily.xls\WCVI$

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.

9-11. **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 (Source: 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%). 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.
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 days).

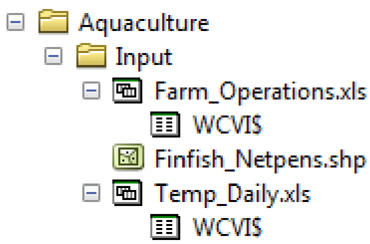
Running the Model

Note about terminology used here: 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.

1. **Exploring the workspace and input folders**. These folders will hold all input, intermediate 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.

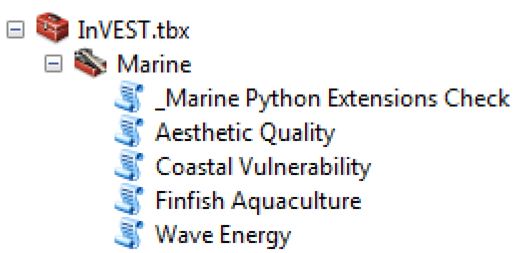
Exploring a project workspace and Input data folder. The *\InVEST\Aquaculture* folder holds the main working folder for the model and all other associated folders. Within the *Aquaculture* folder there will be a subfolder named *'Input'*. This folder holds most of the GIS and tabular data needed to setup and run the model.

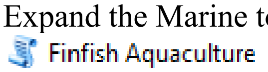
The following image shows the sample folder structure and accompanying GIS data. We recommend using this folder structure as a guide to organize your workspaces and data. Refer the following screenshots below for examples of folder structure and data organization.



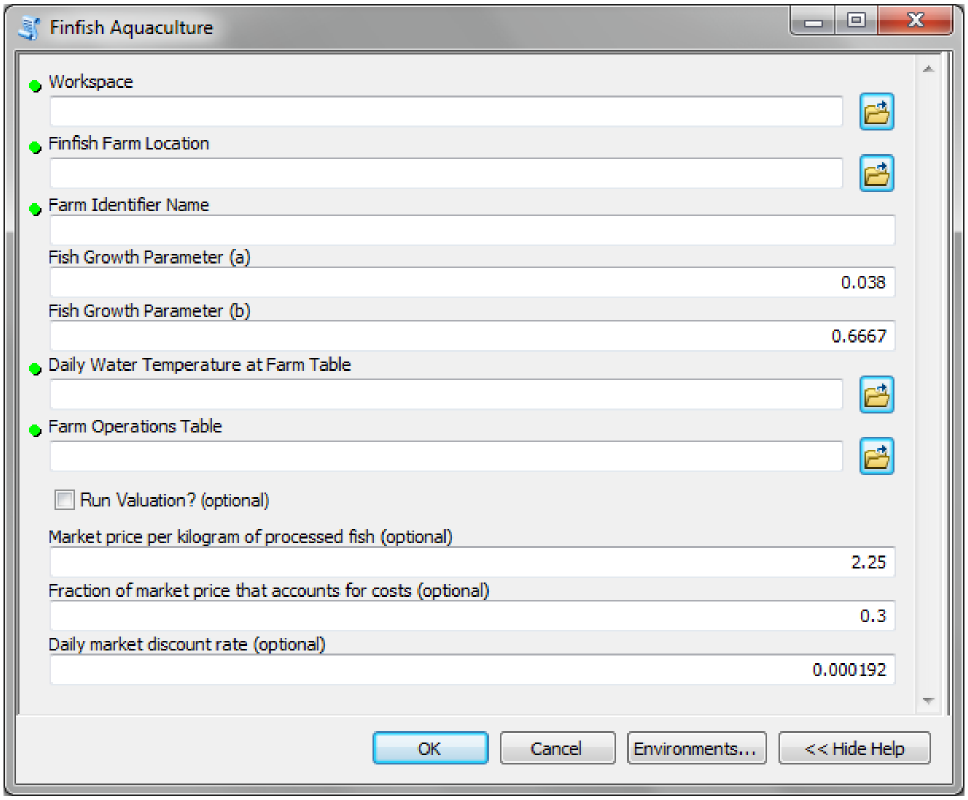
2. **Creating a run of the model.** The following example describes how to set up the Aquaculture model using the sample data provided with the InVEST download. We expect users to have location-specific data to use in place of the sample data. These instructions provide only a guideline on how to specify to ArcGIS the various types of data needed and do not represent any site-specific model parameters. See the **Data Needs** section for a more complete description of the data specified below.

   a. Click the plus symbol next to the InVEST toolbox.

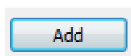


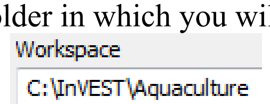
   b. Expand the Marine toolset and click on the Finfish Aquaculture script 

   to open the model.

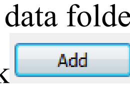


c. Specify the Workspace. Open  the *InVEST* workspace. If you created your own workspace folder (Step 1), then select it here.

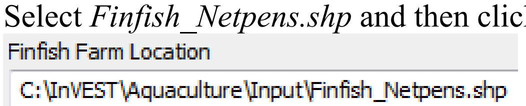
Select the *Aquaculture* folder and click 

to set the main model workspace. This is the folder in which you will find the intermediate and final outputs when the model is run. 

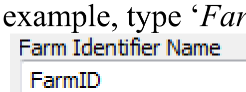
d. Specify the Finfish Farm Location. This represents the geographic area over which the model will be run. This example refers to *Finfish_Netpens.shp* supplied in the sample data.
Open  the *\InVEST\Aquaculture\Input* data folder.

Select *Finfish_Netpens.shp* and then click 

.

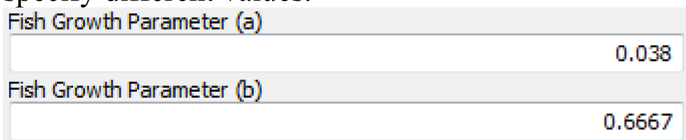


e. Specify the Farm ID Field. The model requires the name of the attribute column from the Finfish Farm Location shapefile that contains the unique farm code. For this example, type '*FarmID*' directly into the text box space.

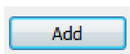


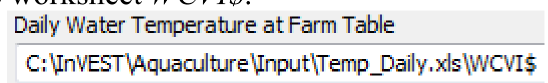
Note: Make sure you enter the Farm Identifier Name text string here exactly as it appears in the Finfish Farm Location shapefile's attribute table heading.

f. Specify the Fish Growth Parameters (a) and (b). These values are the growth parameters required by the model. Default values of 0.038 and 0.6667 (appropriate for Atlantic salmon only) are supplied for you. You can type directly into the text box to specify different values.



g. Specify the Temperature Data. The model requires an Excel table of daily time series of temperature data. Open  the *\InVEST\Aquaculture\Input* data folder. Double left-click on *Temp_Daily.xls* and select the worksheet *WCVI$*.

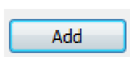
Click 

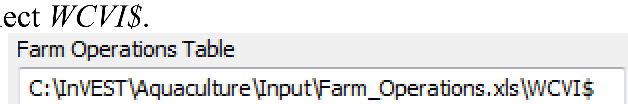
to make the selection. 

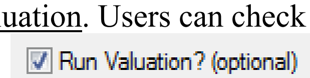
Note: ArcMap and the model will not recognize the Excel sheet as valid data if it is added to the Data View. It is best to add Excel data directly to the model using the Open and Add buttons and navigating to the data.

h. Specify the Farm Operations Data. The model requires an Excel table of farm-specific operation data.

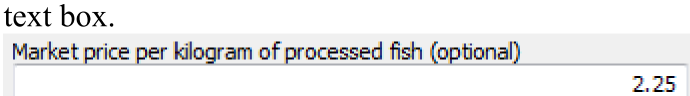
Open  the *\InVEST\Aquaculture\Input* data folder, double left-click *Farm_Operations.xls* and select *WCVI$*.

Click 

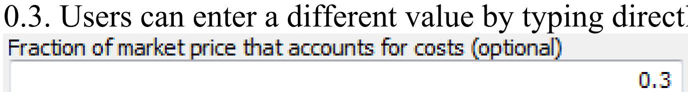
to make the. 

i. Choose whether to run the economic valuation. Users can check the Run Valuation to conduct an economic valuation analysis. 

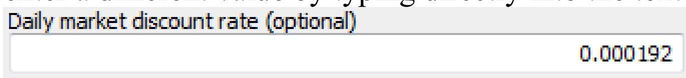
j. Specify the market price of processed fish ($/per kilogram) (Optional). This optional parameter is the market price for a specific processed fish species. The default is given as 2.25 for Atlantic salmon. Users can enter a different value by typing directly into the text box.



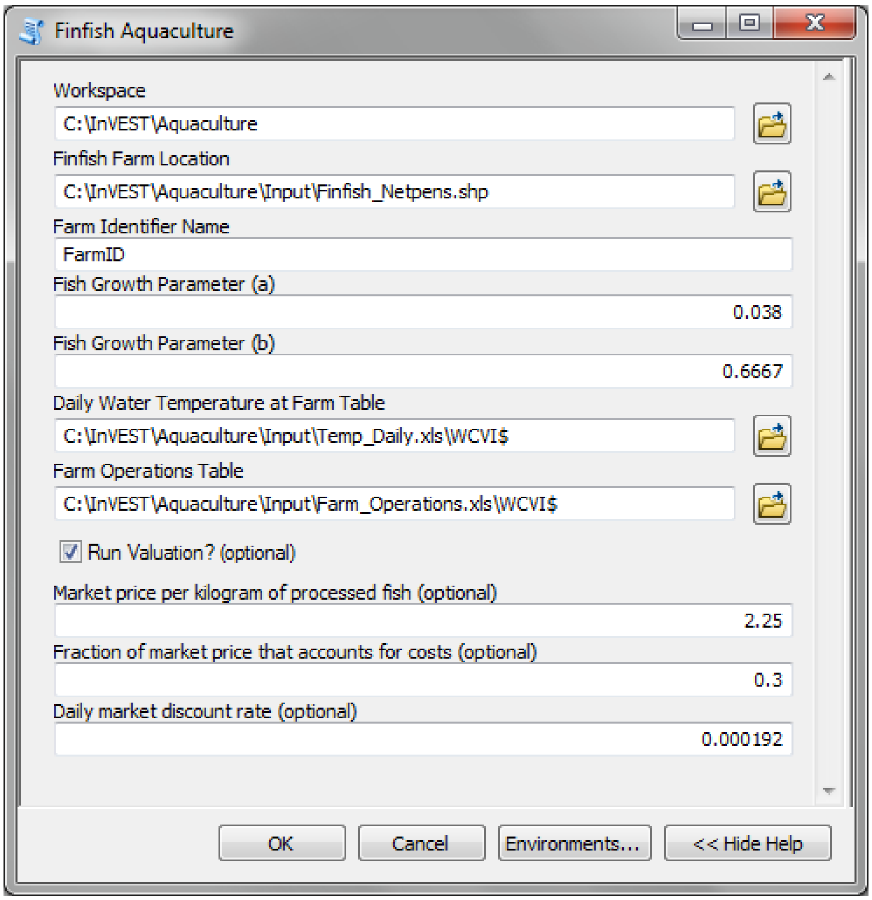
k. Specify the fraction of market prices attributable to costs (Optional). This optional parameter is the fraction of market price attributable to costs. The default is given as 0.3. Users can enter a different value by typing directly into the text box.

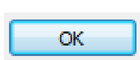


l. Specify the daily market discount rate (Optional). This optional parameter is the discount rate for a type of fish. The default is given as 0.000192 (0.0192%). Users can enter a different value by typing directly into the text box.

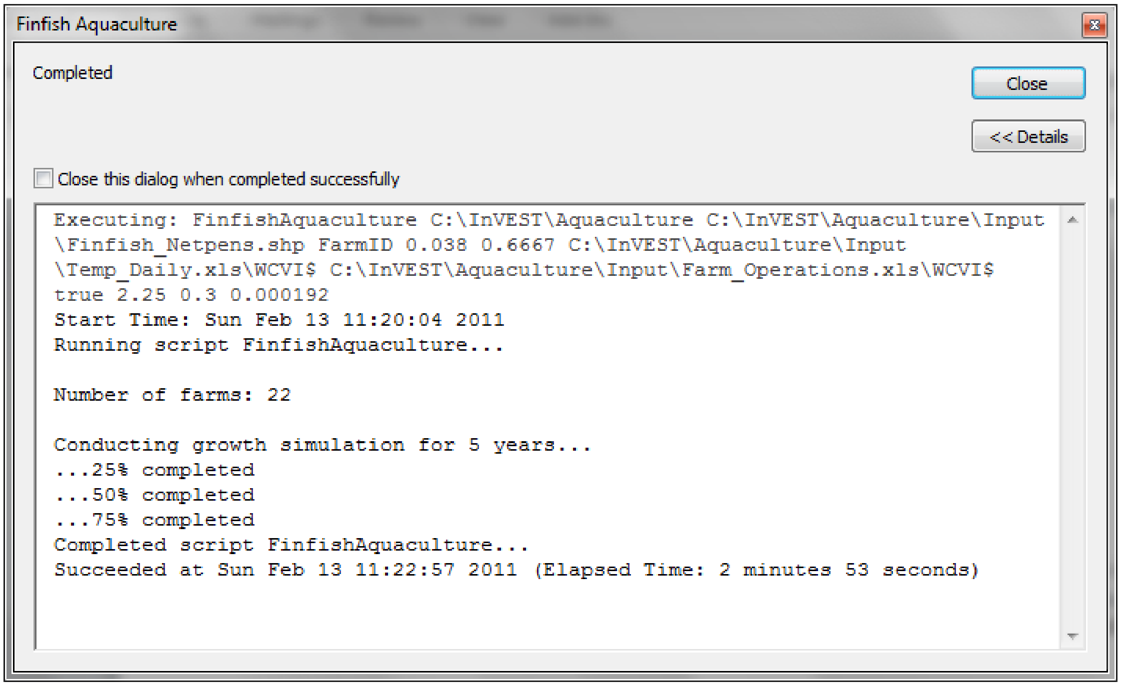


m. At this point the model dialog box is ready for a complete run of the Finfish Aquaculture model.



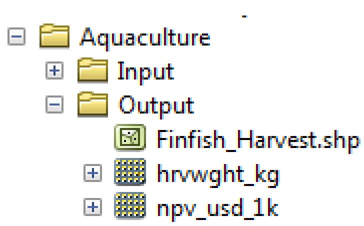
Click 

to start the model. 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 necessary to complete the model run.



3. **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, may contain several types of spatial data, which are described in the **Interpreting Results** section.



The Results.html file located in *\InVEST\Aquaculture\Output* can help you to interpret the model results in terms of fish production and the economic valuation.

You can view the output spatial data in ArcMap (*Finfish_Harvest.shp* shapefile) 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".

Interpreting the Results

Model Outputs

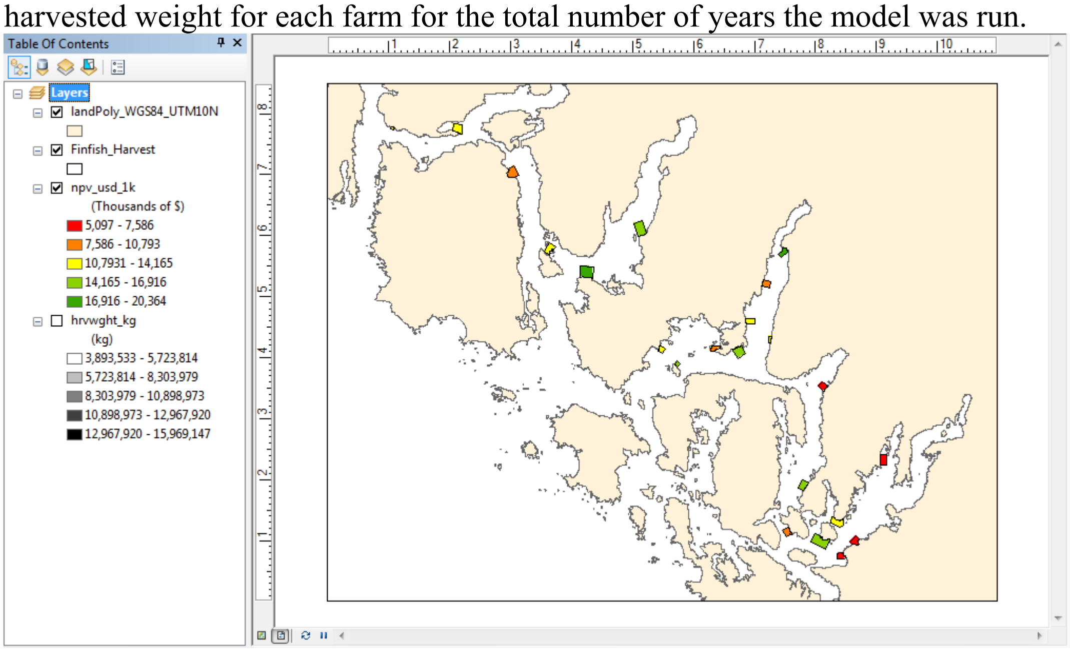
The following is a short description of each of the outputs from the Aquaculture tool. Each of these output files is automatically saved in the "Output" folder that is saved within the user-specified workspace directory:

Final results are found in the output folder of the workspace for this model. The model produces three main output files:

1. 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.

2. Output\hrvwght_kg → A raster file showing total harvested weight in kg for each farm for the total number of years the model was run.

3. Output\npv_usd_1k → A raster file showing total net present value (thousands of $) of the harvested weight for each farm for the total number of years the model was run.



Sample Finfish Aquaculture Model Outputs (#1-3)
(showing net present value for 5 year simulation)

4. Output\ HarvestResults_[date and time].html → An HTML document containing three tables that summarize the inputs and outputs of the model. Cells highlighted in yellow

indicate values that have also been added to the attribute table of the netpens shapefile. Cells highlighted in red should be interpreted as null values since they appear when the valuation option was not selected by the user.

Input:

- **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

Outputs:

- **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

Aquaculture Model (Finfish Harvest)

This page contains results from running the Marine InVEST Finfish Aquaculture model.

Cells highlighted in yellow are values that were also populated in the attribute table of the netpens feature class. Cells highlighted in red should be interpreted as null values since valuation was not selected.

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
5	0.06	5.4	890000	20	10
6	0.06	5.4	900000	60	90
7	0.06	5.4	840000	60	90

Farm Harvesting (output)

Farm ID Number	Cycle Number	Days Since Outplanting Date (including fallowing period)	Harvested Weight (kg/cycle)	Net Revenue (Thousands of $)	Net Present Value (Thousands of $)	Outplant Day (Julian Day)	Outplant Year
5	1	430	3871281	6097	5624	20	2
4	1	515	3174152	4999	4607	60	2
6	1	516	3909531	6157	5673	60	2
20	1	524	2985959	4702	4326	60	2
22	1	524	1514616	2385	2194	60	2
3	1	527	2554012	4022	3698	50	2
12	1	446	1819689	2866	2633	70	2
11	1	534	3210736	5056	4643	60	2

Farm Result Totals (output)

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	13331	4	10403796
2	13331	4	10403796
3	10115	3	7679300
4	12611	3	9529357
5	20082	4	15520044
6	15535	3	11748082
7	14165	3	10898973

Sample Finfish Aquaculture Model HTML Output (#4)
(showing only first few rows of each table)

5. 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.

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.

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.

Aesthetic Quality

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 aesthetic views 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. This is a "Tier 0" model.

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 aesthetic 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 aesthetic 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.

Although this model does not 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.

The Model

The aesthetic 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 tool requires users to provide a DEM and a point shapefile that identifies the locations of sites that contribute to visual impacts. The viewshed analysis is then computed over a user-defined area of interest (AOI) using the ArcGIS viewshed tool.

The model creates three outputs that can be used to assess the visible impact of any type of facility added to the marine environment. The first output is a visual quality raster layer that records the number of sites (e.g. wave energy facilities or aquaculture farms) that are visible from a given raster cell. These counts are then used to classify each raster cell with the following classes: 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 user is provided with an option to choose from two classification schemes (Quartiles or Jenks Natural Breaks) to view the spatially-explicit classified output.

The second 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. You can provide your own population raster data (note that it must be in WGS84 datum).

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.

How it works

The InVEST aesthetic quality model is a set of wrap-around functions that employs ArcGIS's viewshed tool. ArcGIS's viewshed tool implements line of sight computations; the algorithm used by the tool is proprietary to ESRI and there is little documentation of the algorithm details. Users who are interested in further details should consult the ArcGIS online documentation.

Limitations and Simplifications

The global DEM included with the aesthetic 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.

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 and the maps and data tables used by the model. See the FAQ document (http://invest.ecoinformatics.org) 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 without conducting polygon overlap analysis.

1. **Workspace Location (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 "AestheticQuality" 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\AestheticQuality\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). At the start, the model will check that the AOI is a polygon feature and overlaps with the DEM input. If not, it will stop and provide feedback.

   Names: File can be named anything, but no spaces in the name
   File type: polygon shapefile (.shp)
   Sample path: \InVEST\AestheticQuality\AOI_WCVI.shp

3. **Point Features Contributing to Negative Aesthetic Quality (required)**. The user must specify a point feature layer that indicates locations of objects that contribute to negative aesthetic quality, such as aquaculture netpens or wave energy facilities. In order for the viewshed analysis to run correctly, the projection of this input must be consistent with the project of the DEM (input #4). At the start, the model will check that inputs #3 and #4 have consistent projections. If not, it will stop and provide feedback. For instructions on how to create a point shapefile, see the InVEST FAQ document.

   Names: File can be named anything, but no spaces in the name
   File type: point shapefile (.shp)
   Sample path: \InVEST\AestheticQuality\AquaWEM_points.shp

4. **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 #3 are visible.

   Name: File can be named anything, but no spaces in the name and less than 13 characters
   Format: standard GIS raster file (e.g., ESRI GRID or IMG), with elevation values
   Sample data set: \InVEST\AestheticQuality\Base_Data\Marine\DEMs\claybark_dem

5. **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.

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): 0.13

6. **Viewshed Visual Quality Classification Type (required).** This drop down box allows users to specify which type of classification scheme should be used for the model's visual quality output. By selecting "Quartiles", this output will calculate breaks based on the 25th, 50th and 75th percentiles of the distribution. By selecting "Natural Breaks", this output will calculate breaks based on the Jenks Optimal, or Jenks' Natural Breaks algorithm. Natural Breaks aims to present a series of break values that best represent the actual breaks observed in the data. In this way, the actual clustering of data values is preserved.

   Names: 1) Quartiles –or– 2) Natural Breaks
   File type: drop down options
   Sample (default): Quartiles

7. **Global Population Raster (required).** A global raster layer is required to determine population within the AOI's land-seascape where features from input #3 are visible and not visible.

   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

Optional Inputs

The next series of inputs are optional, but may be required depending on other decision inputs.

8. **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.

   Names: A numeric text string (positive integer)
   File type: text string (direct input to the ArcGIS interface)
   Sample (default): 500

9. **Polygon Features for Overlap Analysis (optional)**. The user has the option of providing a polygon feature layer where they would like to determine the impact of points (input #3) on visual quality. This input must be a polygon and have a WGS84 datum. 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 #3.

Names: File can be named anything, but no spaces in the name
File type: polygon shapefile (.shp)
Sample path: \InVEST\AestheticQuality\BC_parks.shp

10. **Projection for Overlap Analysis (optional)**. In order to accurately calculate area and determine overlap, the model must project the polygon feature (input #9) into a projection with meters as the units. As an input the polygons from input #9 can be projected or unprojected, but must have a WGS84 datum. This projection input is required when input # 9 (polygon features for overlap analysis) is specified.

    File type: projection files provided by ArcGIS (.prj)
    Sample path: Coordinate Systems\Projected Coordinate Systems\UTM\WGS 1984\ WGS 1984 UTM Zone 10N.prj

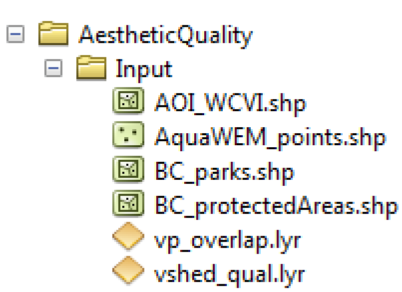
Running the Model

Note about terminology used here: 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.

1. **Exploring the workspace and input folders**. These folders will hold all input, intermediate 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.

    Exploring a project workspace and Input data folder. The *\InVEST\AestheticQuality* folder holds the main working folder for the model and all other associated folders. Within the *AestheticQuality* 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.

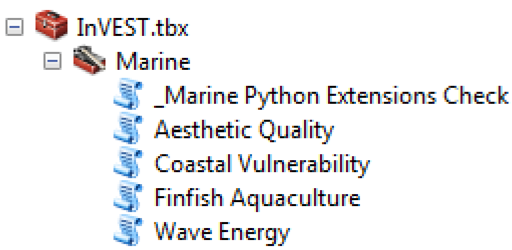
    The following image shows the sample folder structure and accompanying GIS data. We recommend using this folder structure as a guide to organize your workspaces and data. Refer to the following screenshots below for examples of folder structure and data organization.

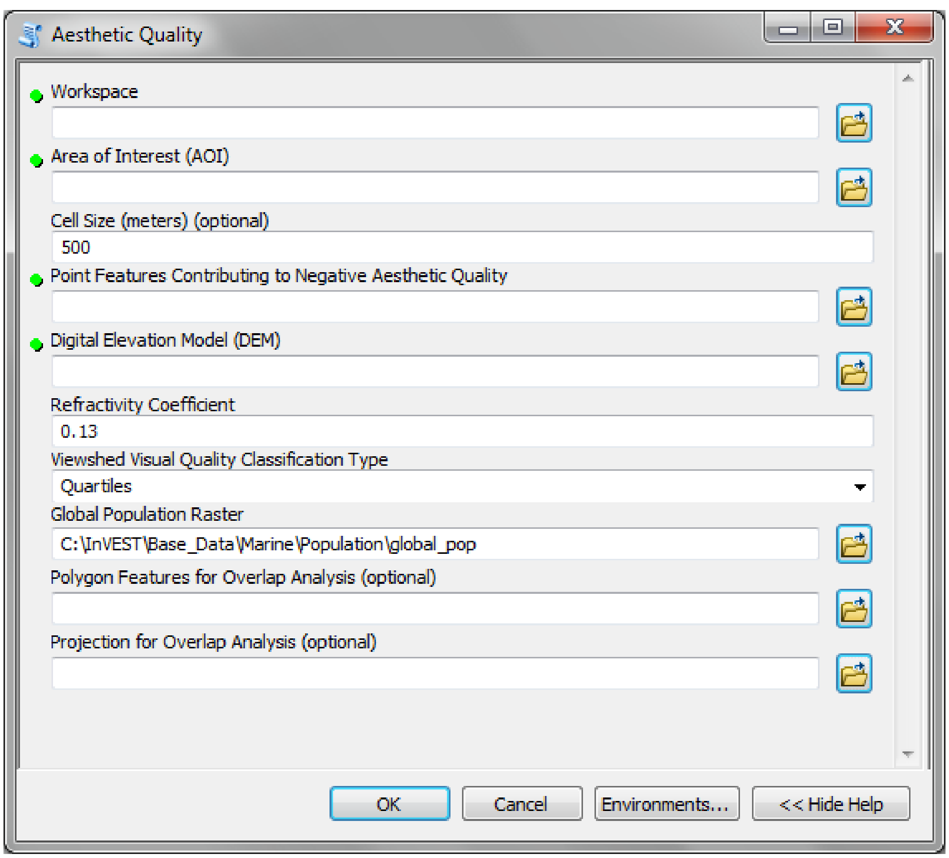
2. **Creating a run of the model.** The following example of setting up the Aesthetic Quality model uses the sample data and folder structure supplied with the InVEST installation

package (see the **Data Needs** section for a more complete description of the data). These instructions only provide a guideline on how to specify to ArcGIS the various types of data needed and does not represent any site-specific model parameters. Users might choose different input parameters and/or have location-specific data to use in place of the sample data.

a. Click the plus symbol next to the InVEST toolbox.



b. Expand the Marine toolset and click on the Aesthetic Quality script Aesthetic Quality to open the model.



c. <u>Specify the Workspace</u>. Open the *InVEST* workspace. If you created your own workspace folder (Step 1), then select it here.

Select the *AestheticQuality* folder and click Add to set the main model workspace. This is the folder in which you will find the intermediate and final outputs when model is run.

Workspace
C:\InVEST\AestheticQuality

d. Specify the Area of Interest (AOI). The AOI is the geographic area over which the model will be run. This example refers to the *AOI_WCVI.shp* shapefile supplied in the sample data. You can create an AOI shapefile by following the **Creating an AOI** instructions in the **FAQ document** (http://invest.ecoinformatics.org).

Open the *\InVEST\AestheticQuality\Input* data folder.

Area of Interest (AOI)
C:\InVEST\AestheticQuality\Input\AOI_WCVI.shp

If you created your own *Input* folder in step 1b, then select it here. Select the AOI shapefile and click Add to make the selection.

e. Specify the Cell Size. This option determines the cell size for the output viewshed raster. The default is "500", meaning the model will run at the 500m resolution utilizing the input DEM. You can type directly into the text box to specify a different value.

Cell Size (meters) (optional)
500

f. Specify the Point Features Contributing to Negative Aesthetic Quality. This vector dataset represents points that have undesirable effects on aesthetic viewing quality.

Open the *Input* data folder *\InVEST\AestheticQuality\Input* and click Add the *AquaWEM_points.shp* shapefile.

Point Features Contributing to Negative Aesthetic Quality
C:\InVEST\AestheticQuality\Input\AquaWEM_points.shp

g. Specify the Digital Elevation Model. The digital elevation model provides the base upon *InVEST\Base_Data\Marine\DEMs* folder, select the *claybark_dem* raster and click Add.

Digital Elevation Model (DEM)
C:\InVEST\Base_Data\Marine\DEMs\claybark_dem

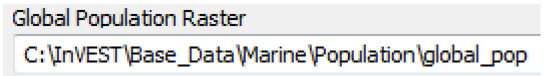
h. Specify the Refractivity Coefficient. The model requires a refractivity coefficient. The default value is value 0.13. You can type directly into the text box to specify a different value.

Refractivity Coefficient
0.13

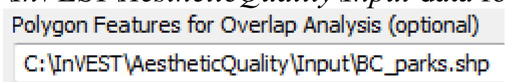
i. Specify the Viewshed Visual Quality Classification Type. This option specifies how cell values in the output viewshed raster are grouped by value classes. Two options are provided: Quartiles and Natural Breaks.

Viewshed Visual Quality Classification Type
Quartiles
Quartiles
Natural Breaks

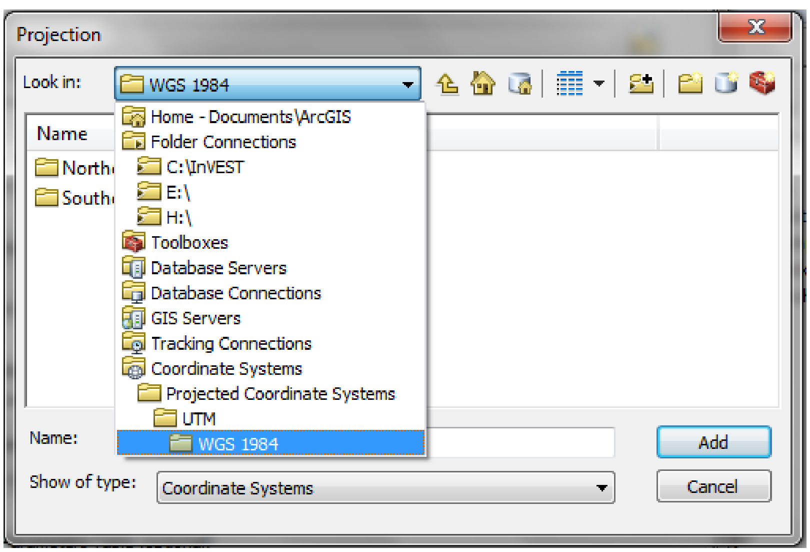
j. <u>Specify Global Population Raster</u>. This dataset represents raster cells of population and is required for the viewshed analysis. Open the *\InVEST\Base_Data\Marine\Population* folder and click the *global_pop* raster.

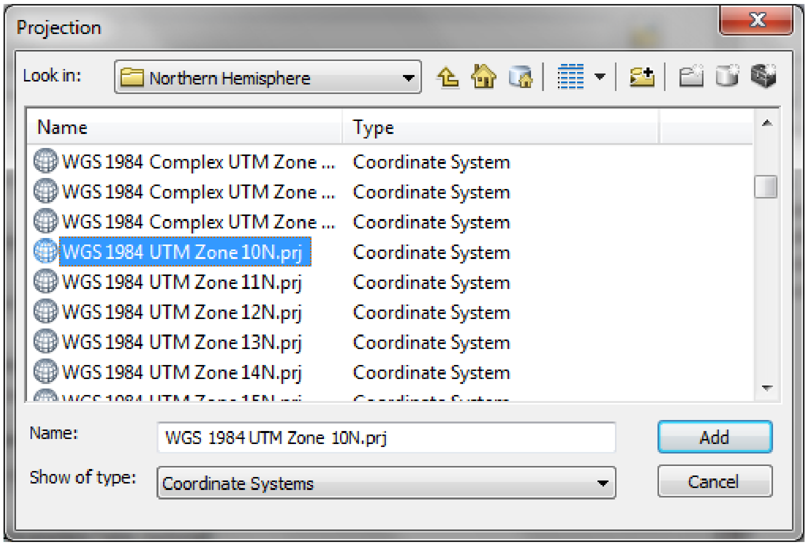


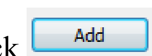
k. <u>Specify Polygon Features for Overlap Analysis (Optional).</u> This vector dataset represents polygon areas to be considered for the viewshed analysis. Open the *\InVEST\AestheticQuality\Input* data folder and add the *BC_parks.shp* shapefile.



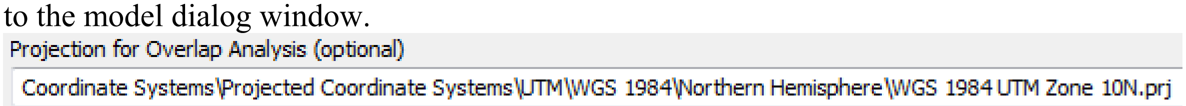
l. <u>Specify Projection for Overlap Analysis (Optional)</u>. The Projection file is specified to set the projection and coordinate information necessary to run the overlay analysis. Open the Coordinate Systems folder near the bottom of the Look In list and path to the *UTM\WGS 1984* folder.





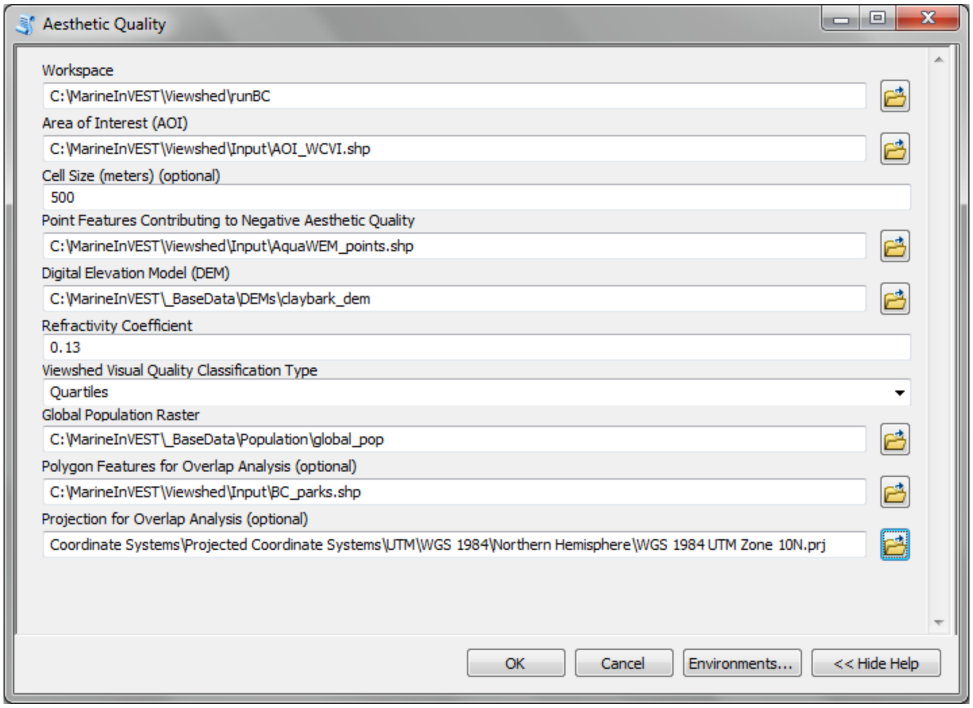
Select the WGS 1984 UTM Zone 10N.prj projection file and click 

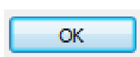
to add it to the model dialog window.



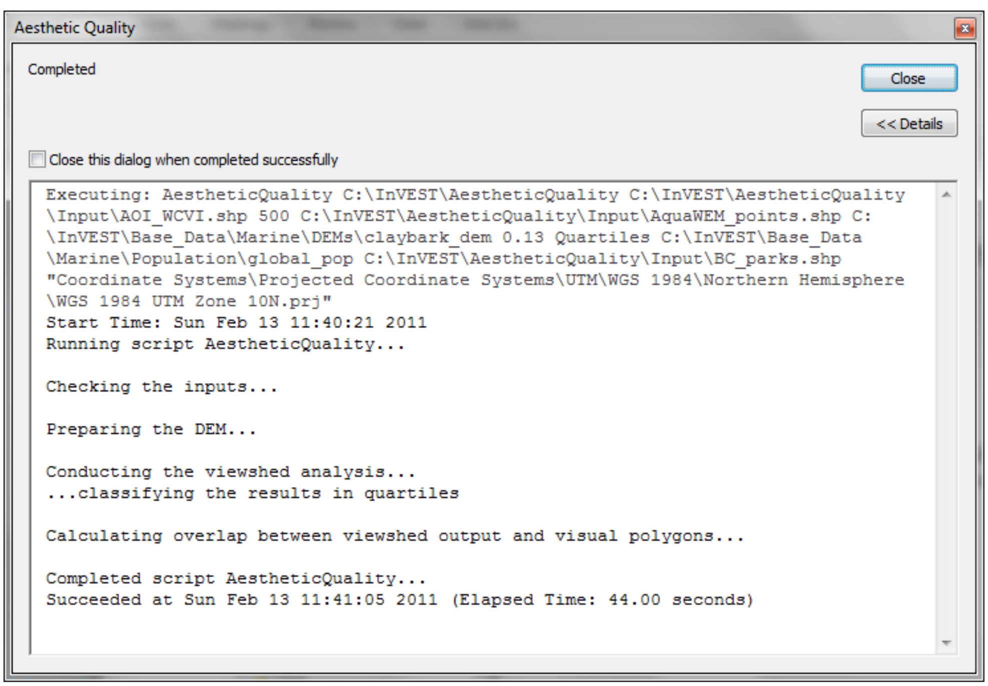
It is assumed that all of your input data are in the same projection and coordinate systems with matching datum. If you need to re-project your data see the Projection section in **Getting Started section** or the **FAQ document** at http://invest.ecoinformatics.org.

m. At this point the model dialog box is completed for a complete run of the Aesthetic Views model.



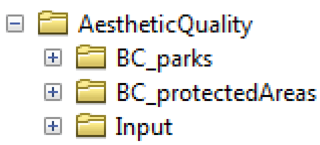
Click 

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 necessary for the model run.



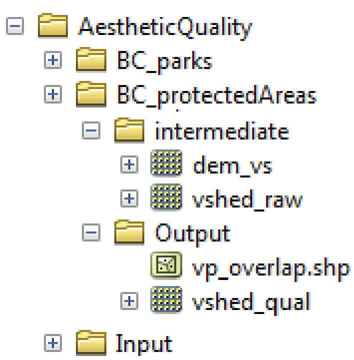
3. Multiple runs of the model

The model setup is the same as for a single run, but the user needs to specify a new workspace for each new run. Make sure each new workspace exists under the main model workspace folder (i.e. *AestheticQuality* 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 Aesthetic Quality model based on two different visual polygon shapefiles (BC_parks.shp and BC_protectedAreas.shp), you could use the Input data folder under main *AestheticQuality* folder and create two new workspace folders, BC_parks and BC_protectedAreas. See below for an example of the folder setup.



4. 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, may 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".

Interpreting Results

Model Outputs

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

<u>"Output" Folder</u>

1. Output\ vshed_qual
   - This raster layer contains a field that classifies (based on either quartiles or natural breaks) 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 easily be symbolized by importing the symbology from the file "\AestheticQuality\Input\vshed_qual.lyr"

2. 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 aesthetic 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 "\AestheticQuality \Input\vp_overlap.lyr"

3. Output\ populationStats_[date and time].html
   a. 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 aesthetic quality are visible) and 2) affected (one or more sites visible).

4. 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.

<u>"intermediate" Folder</u>

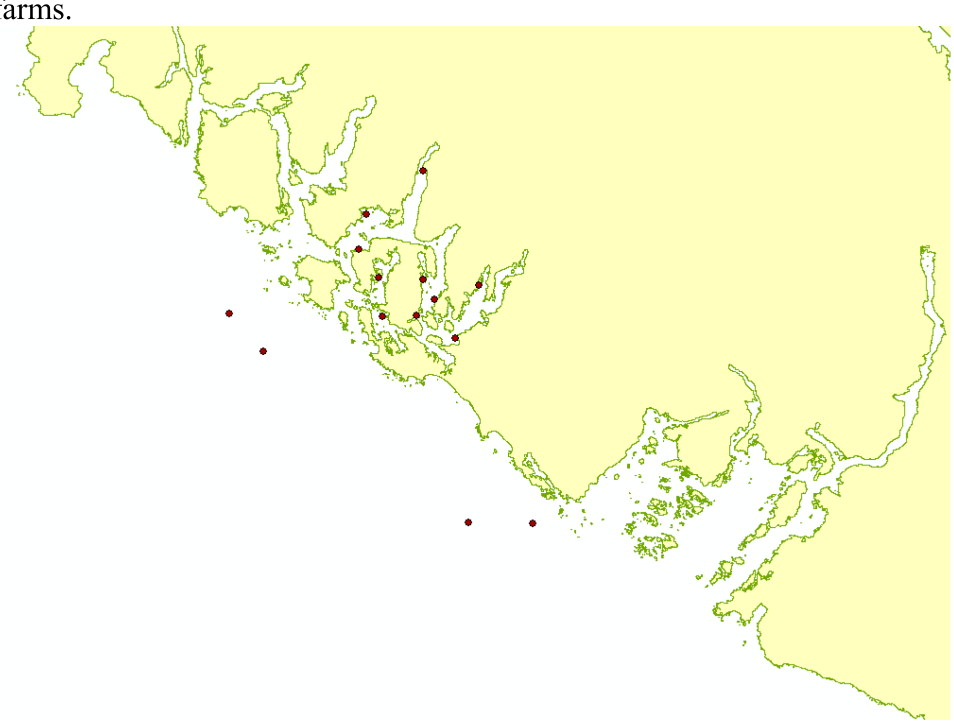
1. intermediate\dem_vs
   - This raster layer is the modified DEM within the user-specified extent. The portions of the DEM that are below sea-level are converted to a value of "0" since all viewing on the ocean will be at the surface.

2. intermediate\vshed_raw
   - 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 contributing to negative

aesthetic quality.   For example, all cells with a value of "4" would indicate that at that location four points are visible.

Case Example Illustrating Results

The following example illustrates the aesthetic views model. In this example, we examine the visual footprint resulting from potential wave energy facilities and aquaculture farms. The following figures and maps are for example only, and are not necessarily an accurate depiction of WCVI. In the first figure, we show the locations of the sites of potential wave energy facilities and aquaculture farms.



In this example, there are four offshore wave energy facilities and ten aquaculture facilities. We then run the aesthetic views model to determine the visual footprint of these potential facilities. To run the model, we first create an area of interest polygon that encompasses all of the site locations and the portion of the sea and landscape that we are interested in evaluating. We then apply an upper bound of 8 km on the search radius. This limits the search distance to 8 km when identifying areas that are visible from each observation point. This upper bound is applied by adding the field RADIUS2 to the shapefile specifying the point features contributing to negative aesthetic quality. To limit the search to 8 km, each point is assigned a value of -8000 as shown in the following figure.

Table

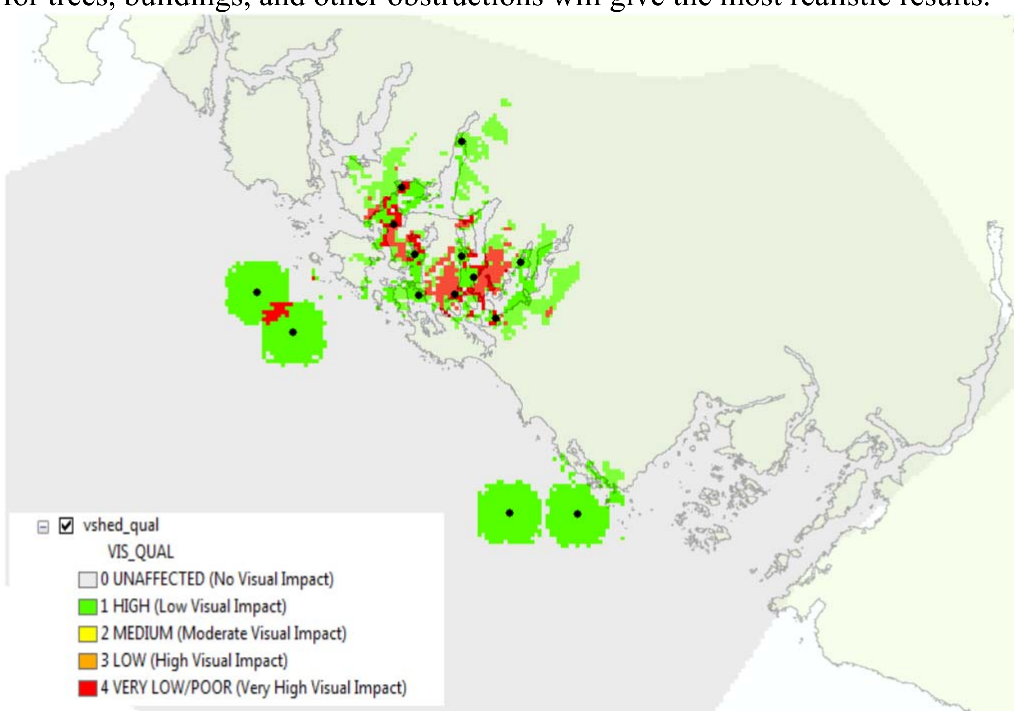
AquaWEM_Real

FID	Shape *	Name	RADIUS2
0	Point		-8000
1	Point		-8000
2	Point		-8000
3	Point		-8000
4	Point		-8000
5	Point		-8000
6	Point		-8000
7	Point		-8000
8	Point		-8000
9	Point		-8000
10	Point		-8000
11	Point		-8000
12	Point		-8000
13	Point		-8000

After completing the steps outlined in the “Running the model” section, we obtain the following map that classifies the visual impacts of these sites.

Classification of visual quality

The resulting map shows the footprint of visual quality from offshore wave energy sites and the aquaculture facilities. The cells highlighted in red are the areas with the highest visual impact; the cells highlighted in green have the lowest visual impact. The grey cells have no visual impact. It is clear from the visual quality map that most offshore areas experience low visual impacts from the wave energy facilities, whereas areas surrounding the clustered aquaculture facilities experience the highest visual impacts. Please be aware that the quality of the viewshed model results depends on the quality of the DEM used in the analysis. Fine resolution DEMs that account for trees, buildings, and other obstructions will give the most realistic results.



Resident population impacted by visual disamenities
In addition to producing a map of the visual footprint of objects located offshore, the aesthetic quality model also provides a count of the resident population that falls within this visual footprint. The viewshed model uses the Gridded Rural-Urban Population Model Project (GRUMP) dataset to extract the population counts within grid cells that are visible from any of the offshore sites. These counts are then tabulated and documented in the “PopulationStats.html” file found in the output folder. For this example, the number of residents unaffected by the offshore sites is 8554

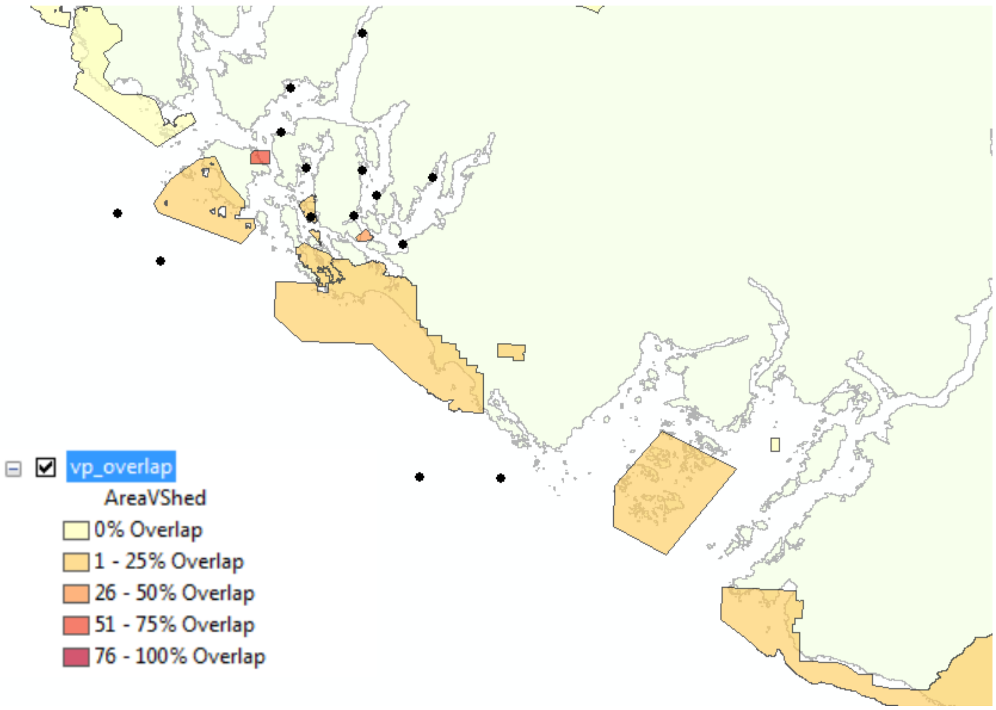
Population Statistics

Number of Visible Sites	Population
0 (unaffected)	8554
1 or more sites visible	3735

and the population count that falls within grid cells that can see at least one offshore site is 3735. Users again should be reminded that the GRUMP dataset is based on site-specific census data and may not accurately reflect the actual population that uses a particular area. This is particularly true for areas important for tourism and other seasonal activities that census data will not account for.

Viewshed overlap with protected areas
The final optional output of the aesthetic quality tool uses a set of user-specified polygons and computes the percent area within each polygon from which at least one offshore site is visible. To illustrate these results, we use a set of polygons that represent protected areas in the same study area explored above.

The protected areas are shown in the above figure as green polygons and the points represent the location of the offshore wave energy facilities and aquaculture sites. For each protected area in the user-specified area of interest, the model then computes the percentage of each protected area that falls within the viewshed of the wave energy and aquaculture sites. The figure below shows the results for a selection of the protected areas included in the example.



From this example, we see that for most of the protected areas, 1 – 25% of their total area falls within the viewshed footprint of the wave energy and aquaculture sites. For one of the smaller protected areas, 51 – 75% of its area falls within the viewshed footprint. These results are not spatially explicit at a fine scale because they do not indicate the exact locations from which one could see the facilities. However, these locations can be identified from the previous aesthetic quality results.

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.

BIODIVERSITY: HABITAT QUALITY & RARITY

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.

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 and Rarity 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.

The Model

The InVEST biodiversity model (Tier 1) combines information on LULC and threats to biodiversity to produce habitat quality and rarity 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-à-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 environment 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 (H_{prarie}) equals 1), but will also make use of a managed hayfield or pasture in a pinch (the habitat score for the LULC hayfield ($H_{hayfield}$) and pasture ($H_{pasture}$) 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 $r = 1, 2, \ldots, R$ 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 km^2). 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,

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

$$i_{rxy} = \exp\left(-\left(\frac{2.99}{d_{r\max}}\right)d_{xy}\right) \text{ if exponential} \qquad (2)$$

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 r_y's disturbance zone. (If the expontential funcion 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.) All threats must have a maximum threat distance > 0. To reiterate, if we have assigned species group-specific habitat suitability scores to each LULC then threat impact over spece should be specific to the modeled species group.

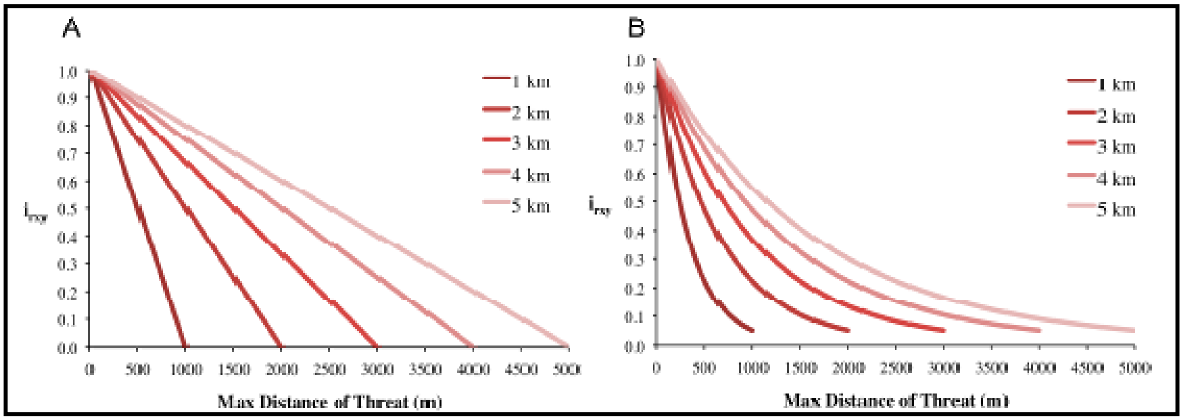


Figure 1. *An example of the relationship between the distance-decay rate of a threat and the maximum effective distance of a threat under A) linear and B) exponential.*

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? 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 β_x 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.

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.

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

Adapted from Czech et al. 2000 – we need to get permission if we decide to include

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},

$$D_{xj} = \sum_{r=1}^{R} \sum_{y=1}^{Y_r} \left(w_r \Big/ \sum_{r=1}^{R} w_r \right) r_y i_{rxy} \beta_x S_{jr} \tag{3}$$

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,

$$Q_{xj} = H_j \left(1 - \left(\frac{D_{xj}^z}{D_{xj}^z + k^z} \right) \right) \tag{4}$$

and z (we hard code $z = 2.5$) and k are scaling parameters (or constants). Q_{xj} is equal to 0 if $H_j = 0$. Q_{xj} increases in H_j 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 $1 - \left(\frac{D_{xj}^z}{D_{xj}^z + k^z} \right) = 0.5$. 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 $k = 30$ but the user can change it (see note in Data Needs section, #8). If you are analyzing the impact of various LULC scenarios on a landscape, use the same k for all alternative scenarios. Similarly, use the same spatial resolution for all LULC scenario analyses from the same landscape. If you decide to change your choice of k or the spatial resolution then you must change these parameters for all model runs 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-à-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.

$$R_j = 1 - \frac{N_j}{N_{j,baseline}} \qquad (5)$$

where N_j is the number of grid cells of LULC *j* on the current or projected map and $N_{j,baseline}$ 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:

$$R_x = \sum_{x=1}^{X} \sigma_{xj} R_j \qquad (6)$$

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.

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 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\Base_Data**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.

   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\Base_data**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.

   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\Base_data**lc_samp_bse_b**

4. **Threat data (required).** A 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: *.dbf or *.xls if using ArcGIS 9.3
Rows: each row is a degradation source
Columns: each column contains a different attribute of each degradation source, and must be named as follows:

- **a.** THREAT: the name of the specific threat. **Threat names must not exceed 8 characters.**
- **b.** 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. A threat must have a MAX_DIST > 0.
- **c.** 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.
- **d.** DECAY: Indicates whether the impact of the threat decreases linearly or exponentially across space. Value can be either 0 or 1. A value of 1 indicates a linear decline in impact, while 0 indicates an exponential decline.

Sample Data Set: \Invest\Biodiversity\Input**threats_samp.dbf**

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	1
Paved_rd	4	0.4	1
Agric	8	1	0

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. 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*", "*_c*", or "*_f*" 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\Biodiversity\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 and rarity 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:
   a. ID: unique identifying code for each polygon. FID also works.

b. Access: values between 0 and 1 for each parcel, as described above.
<u>Sample data set:</u> \Invest**access_samp.shp**

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

<u>Name</u>: file can be named anything
<u>File type</u>: *.dbf or *.xls if using ArcMAP 9.3
<u>Rows</u>: each row is a LULC type.
<u>Columns</u>: columns contain data on land use types and their sensitivities to threatss. Columns must be named according to the naming conventions below.

a. 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.
b. NAME: the name of each LULC
c. HABITAT: Each LULC is assigned a habitat score, H_j, 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).
d. 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**.

<u>Sample data set</u>: \Invest\Biodiversity\Input**sensitivity_samp.dbf**

<u>Example</u>: 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 30 but can be set equal to any positive integer. In general, it is best to set *k* equal to half of the highest grid cell degradation value on the landscape. To perform this model calibration run the model once to find the highest degradation value on the landscape and then set *k* to half this value for your landscape (if you plan on evaluating multiple LULC scenarios use the LULC scenario designated as the "current' landscape to calibrate *k*). For example, if a preliminary run of the model generates a degradation map where the highest grid-cell degradation level is 10 then setting *k* at 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 *k* value for all runs that involve the same landscape.

Running the Biodiversity Model

Before running the Biodiversity Model, first make sure that the InVEST toolbox has been added to your ArcMap document, as described in the **Getting Started** chapter of this manual. Second, make sure that you have prepared the required input data files according to the specifications in *Data Needs*. Specifically, you will need (1) a current LULC raster file showing the location of different LULC types in the landscape; (2) a future LULC raster if you wish to project future habitat quality and rarity across the landscape; (3) a baseline LULC map if you wish to express habitat rarity on the current and future landscapes or measure habitat extent and quality on the baseline landscape; (4) a threat data table denoting the intensity and distance over which a degradation source occurs; (5) grids showing the spatial distribution of each threat on each submitted map (current, future, and baseline); (6) a shapefile indicating the relatively accessibility to an area based on protection; (7) a table indicating the habitat suitability for each LULC and the sensitivity of each habitat type to each threat; and (8) a numeric value indicating the half-saturation constant.

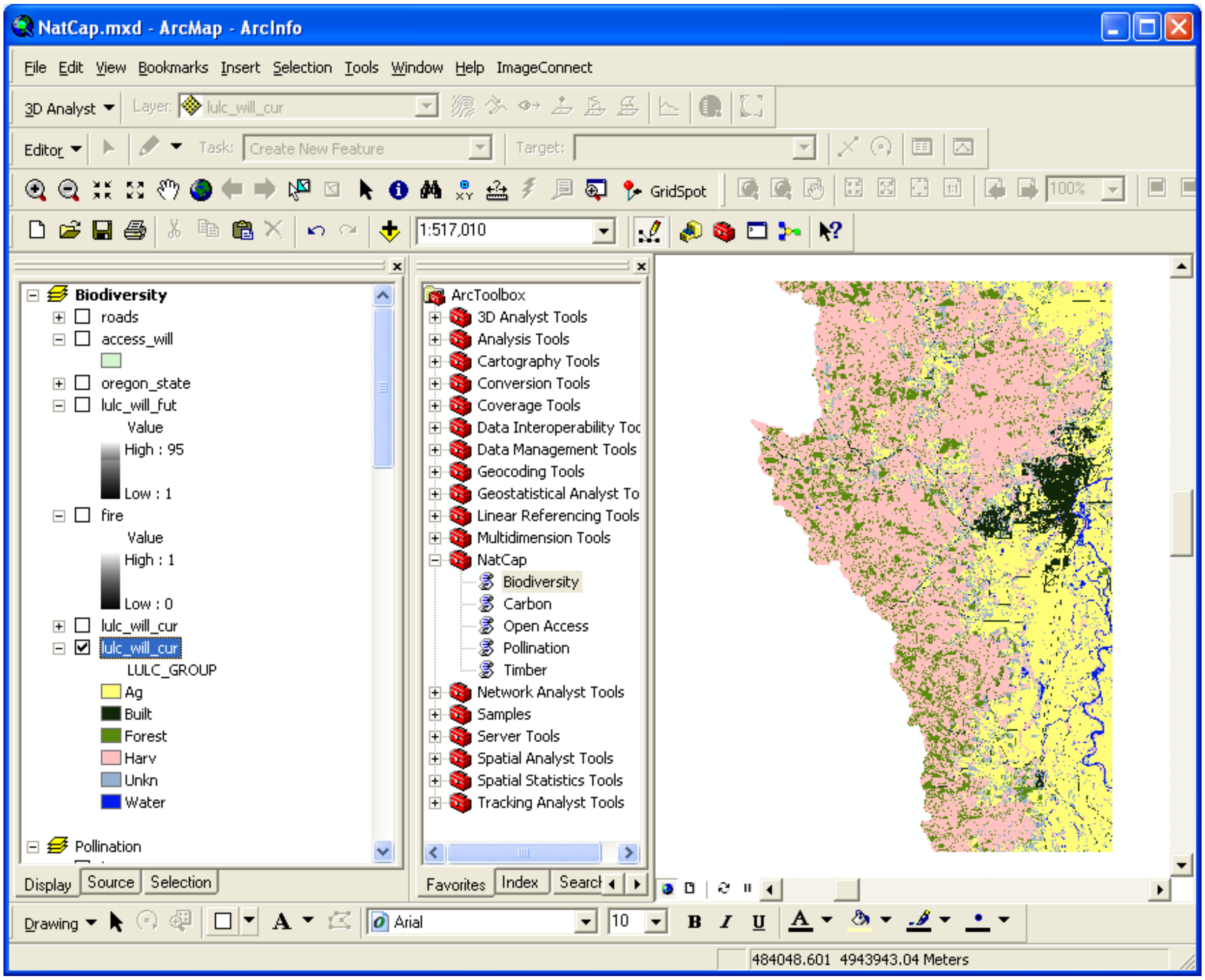
- Create a workspace

You must create a folder in your workspace called "input" and place all your input files here, including all your threat maps. If this is your first time using InVEST and you wish to use sample data, you can use the data provided in InVEST-Setup.exe. If you unzipped the InVEST files to your C-drive (as described in the **Getting Started** chapter), you should see a folder called /Invest/**biodiversity**. This folder should be your workspace. The input files are in a folder called /Invest/biodiversity/**input** and in /Invest/**base_data**.

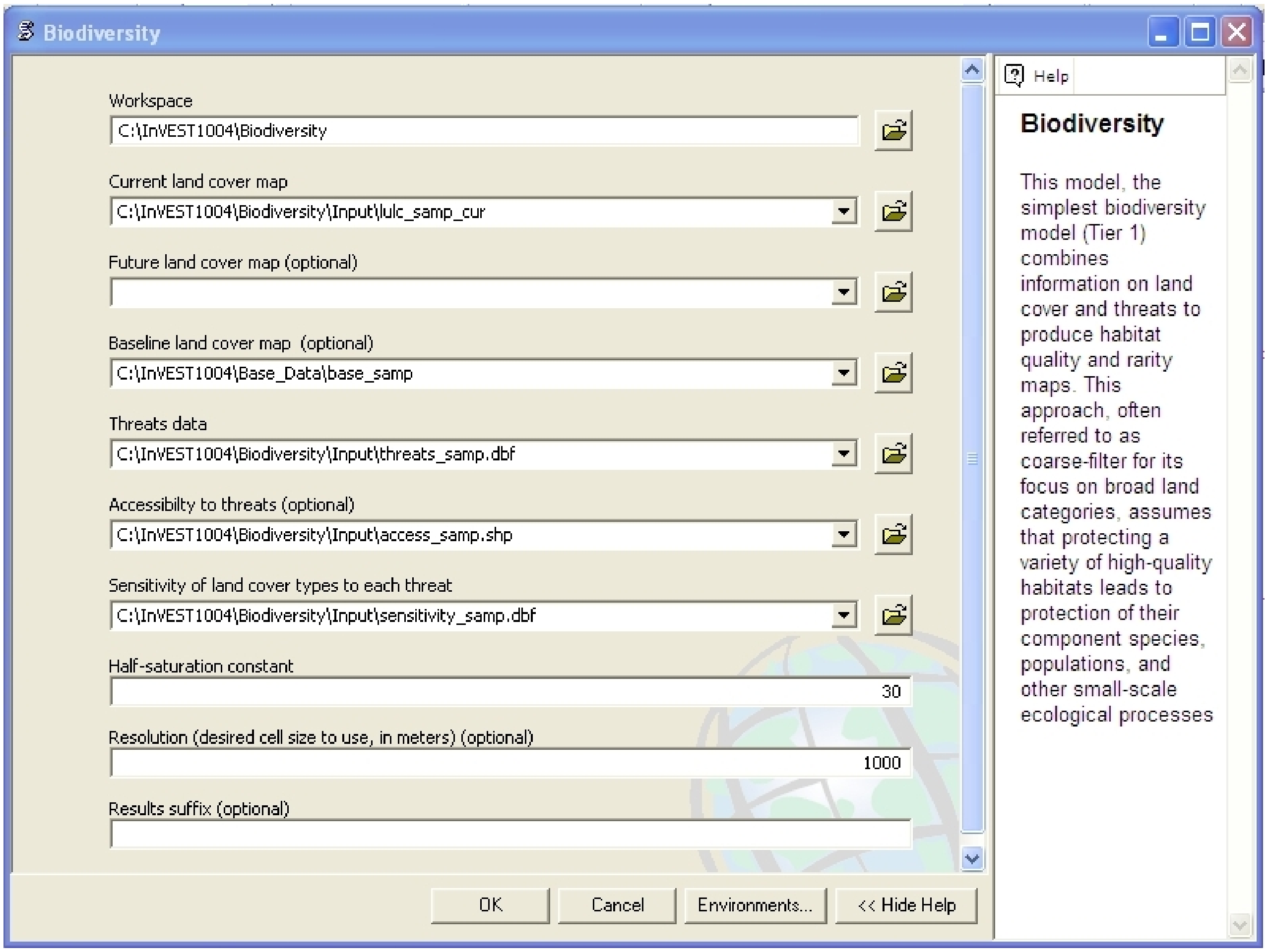
- Open an ArcMap document to run your model.

- Find the InVEST toolbox in ArcToolbox. ArcToolbox should be open in ArcMap, but if it is not, click on the ArcToolbox symbol. See the **Getting Started** chapter if you do not see the InVEST toolbox.

- Click once on the plus sign on the left side of the INVEST toolbox to see the list of tools expand. Double-click on Biodiversity.



- An interface will pop up like the one above that indicates default file names, but you can use the file buttons to browse to your data. When you place your cursor in each space, you can read a description of the data requirements in the right side of the interface. In addition, refer to the ***Data Needs*** section above for information on data formats.



- Fill in data file names and values for all required prompts. Unless the space is indicated as optional, it requires you to enter some data.

- After entering all values as required, click on OK. The script will run, and its progress will be indicated by a "Progress dialogue."

- Upon successful completion of the model, you will see new folders in your workspace called "intermediate" and "output." These folders contain several raster grids which are described in the next section.

- Load the output grids into 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 file's appearance.

- You can also view the attribute data of output files by right clicking on a layer and selecting Open Attribute Table.

Interpreting Results

Parameter Log

Each time the model is run, a text 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.

Final Results

Final results are found in the "Output" folder of the workspace for this module.

degrad_cur[suffix] – 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 $H_j = 0$) get a degradation score of 0. This is a mapping of degradation scores calculated with equation (3).

qual_cur [suffix] – Habitat quality on the current landscape. Higher numbers indicate better habitat quality vis-à-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[suffix] – Relative habitat rarity on the current landscape vis-à-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 R_x (see equation (6)). The rarer the habitat type in a grid cell is vis-à-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[suffix]* and *qual_fut[suffix]* 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[suffix]* 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[suffix]*AND *qual_bse[suffix]* in the output folder.

Recall, if you are setting H_j 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.

<u>Modifying output and creating a landscape biodiversity score</u>
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[suffix]* (if available), *qual_cur[suffix]*, and *qual_fut[suffix]* (if available) maps and then compare scores. Or you could run this model over multiple future LULC maps (re-running the model for each future map) and

compare the landscape-level habitat quality scores of the various alternative future scenarios. Please recall that comparisons of landscape-level habitat quality scores are only valid if the same k value (see Data Need # 8), study area grain and extent, and input tables are used across all model runs.

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 Tier 1 biodiversity model. In this case we would determine 18 aggregate habitat quality scores, once for each modeled species under each scenario. Let G_{s_cur} indicate the set (and number) 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,

$$Q_{s\_cur} = \frac{\sum_{x=1}^{G_{s\_cur}} Q_{xj\_cur}}{G_{s\_cur}} \tag{9}$$

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,

$$Q_{cur} = \left(\sum_{s=1}^{S} Q_{s\_cur}\right) / 9 \tag{10}$$

Then we would repeat for the future landscape with the grid cells in set G_{s_fut} for each species s and the set of Q_{xj_fut}.

Example application of the Tier 1 Habitat Quality model

Here we provide an example of an application of the habitat quality model to Minnesota, U.S.A. (see Polasky et al. 2011 for details). For this study the researchers examined tradeoffs in the provision of a suite of ecosystem services, biodiversity conservation, and economic returns to landowners across several land-use change scenarios. Habitat quality served as a proxy measure for biodiversity conservation. Here they chose to focus on functional diversity within a single taxa for two groups: 1) forest breeding birds and 2) grassland breeding birds (based on Ehrlich et al. 1998). Both groups represent species undergoing widespread population declines in Minnesota (Sauer et al. 2008, Minnesota Audubon 2010).

Using the 1992-2001 National Land Cover Dataset (NLCD) change product (Fry 2009) the authors derived a “current” LULC map (1992) and a “future” LULC map (2001). Because the model was being used to evaluate habitat quality across the state for two separate functional groups, each LULC was assigned two habitat quality scores, one for forest breeding birds and the other for grassland breeding birds. Habitat scores ranged from 0 to 1, with non-habitat scored as 0 and most suitable habitat scored as 1, and marginal habitat scored in between (Table 2 and 3 for grassland and forest birds, respectively). For example, grassland songbirds may prefer native prairie habitat above all other habitat types (habitat suitability = 1), but will also make use of a managed hayfield (habitat suitability = 0.5). Habitat suitability was defined based on conversations with local wildlife experts and relevant literature (e.g., Ehrlich et al. 1988).

Next sources of degradation, or threats, were identified as those human modified LULC types (e.g., urban, agriculture, and roads) that cause habitat fragmentation or edge effects (McKinney 2002, Forman 2003). Edge effects refer to changes in the biological and physical conditions that occur at a patch boundary and within adjacent patches (e.g., facilitating entry of predators, competitors, invasive species, toxic chemicals and other pollutants). The same threat data was used across both functional species groups (Table 4). The sensitivity of a particular species’ habitat type to degradation is based on general principles of landscape ecology and conservation biology (e.g., Forman 1995; Lindenmayer et al. 2008) with a specific focus on the habitat needs of grassland or forest songbirds (e.g., Ehrlich et al. 1998, Mayer and Cameron 2003, Donnelly and Marzluff 2004, Cunningham and Johnson 2006). Sensitivity weights for grassland and forest bird habitat were determined from the literature and expert knowledge (Table 2 and 3, respectively). Here we assume habitat sensitivity to threats is similar across functional groups. Depending on the conservation objective in question habitat sensitivity could vary. Finally for both the current and future habitat quality output maps we used ArcGIS zonal statistics to calculate statewide landscape scores.

Table 2. Sensitivity to threat sources and habitat suitability weights for each LULC type parameterized for breeding grassland songbirds.

LULC	LANDCOVER	L_CCROP	L_URBAN	L_ROADP	L_ROADS	L_ROADT	HABITAT
1	Open water	0.00	0.00	0.00	0.00	0.00	0.00
2	Urban	0.00	0.00	0.00	0.00	0.00	0.00
3	Barren	0.00	0.00	0.00	0.00	0.00	0.00
4	Forest – private ownership*	0.70	0.80	0.80	0.60	0.40	0.00
5	Grassland – private ownership*	0.50	0.70	0.50	0.40	0.30	0.90
6	Agriculture – private ownership*	0.00	0.50	0.50	0.40	0.30	0.30
7	Wetland – private ownership*	0.60	0.80	0.80	0.60	0.40	0.40
40	Forest – private ownership w/ easement*	0.70	0.80	0.80	0.60	0.40	0.00
50	Grassland – private ownership w/ easement*	0.50	0.70	0.50	0.40	0.30	0.95
60	Agriculture – private ownership w/ easement*	0.00	0.50	0.50	0.40	0.30	0.50
70	Wetland – private ownership w/ easement*	0.60	0.80	0.80	0.60	0.40	0.20
400	Forest – public ownership*	0.70	0.80	0.80	0.60	0.40	0.00
500	Grassland – public ownership*	0.50	0.70	0.50	0.40	0.30	1.00
600	Agriculture – public ownership*	0.00	0.50	0.50	0.40	0.30	0.50
700	Wetland – public ownership*	0.60	0.80	0.80	0.60	0.40	0.20

Note: The asterisks denote natural lands that are managed for a variety of economic, environmental, and recreational uses. We subdivided forest, grassland, agricultural, and wetlands based on conservation management codes from the GAP Stewardship database containing land ownership information for the entire state of Minnesota (MNDNR 2000). CCROP = row crop; URBAN = urban development; ROADP = primary roads; ROADS - secondary roads; ROADT = tertiary roads.

Table 3. Sensitivity to threat sources and habitat suitability weights for each LULC type parameterized for breeding forest interior songbirds.

LULC	LANDCOVER	L_CCROP	L_URBAN	L_ROADP	L_ROADS	L_ROADT	HABITAT
1	Open water	0.00	0.00	0.00	0.00	0.00	0.00
2	Urban	0.00	0.00	0.00	0.00	0.00	0.00
3	Barren	0.00	0.00	0.00	0.00	0.00	0.00
4	Forest – private ownership*	0.70	0.80	0.80	0.60	0.40	0.90
5	Grassland – private ownership*	0.50	0.70	0.70	0.50	0.30	0.10
6	Agriculture – private ownership*	0.00	0.50	0.50	0.40	0.30	0.05
7	Wetland – private ownership*	0.50	0.80	0.80	0.60	0.40	0.50
40	Forest – private ownership w/ easement*	0.70	0.80	0.80	0.60	0.40	0.95
50	Grassland – private ownership w/ easement*	0.60	0.70	0.70	0.50	0.30	0.10
60	Agriculture – private ownership w/ easement*	0.00	0.50	0.50	0.40	0.30	0.05
70	Wetland – private ownership w/ easement*	0.60	0.80	0.80	0.60	0.40	0.50
400	Forest – public ownership*	0.70	0.80	0.80	0.60	0.40	1.00
500	Grassland – public ownership*	0.50	0.70	0.70	0.50	0.30	0.10
600	Agriculture – public ownership*	0.00	0.50	0.50	0.40	0.30	0.05
700	Wetland – public ownership*	0.60	0.80	0.80	0.60	0.40	0.50

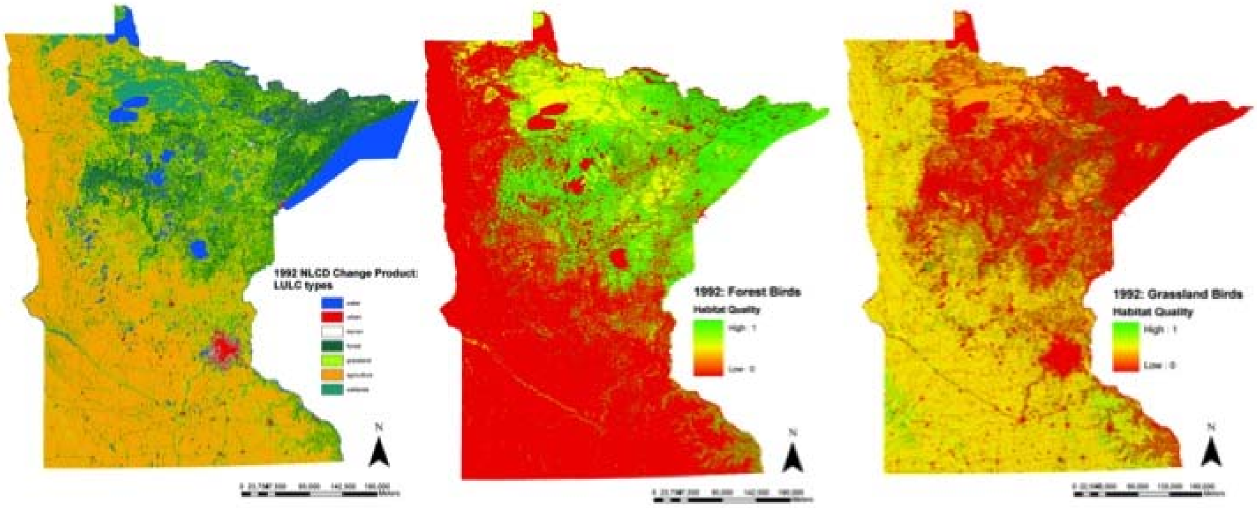
Note: The asterisks denote natural lands that are managed for a variety of economic, environmental, and recreational uses. We subdivided forest, grassland, agricultural, and wetlands based on conservation management codes from the GAP Stewardship database containing land ownership information for the entire state of Minnesota (MNDNR 2000). CCROP = row crop; URBAN = urban development; ROADP = primary roads; ROADS - secondary roads; ROADT = tertiary roads.

Table 4. Weights and effective distances for threat sources.

THREAT	MAX_DIST	WEIGHT	DECAY
ccrop	4.0	0.8	0
urban	5.0	1.0	0
roadp	3.0	0.8	0
roads	2.0	0.7	0
roadt	1.0	0.5	0

Note: CCROP = row crop; URBAN = urban development; ROADP = primary roads; ROADS - secondary roads; ROADT = tertiary roads.

A map of habitat quality scores was generated for each LULC map (1992 and 2001) and functional group combination (there are 4 combinations). In Figure 2 maps of habitat quality scores in 1992 for both functional groups are presented. In general, habitat quality scores for forest breeding birds was greater in the northeastern portion of the state while habitat quality for grassland breeding birds as greatest in the west and southwestern portion of the state. The overall change in the habitat score between 1992 and 2001 due to LULC change is summarized in the table in Figure 2. The state-wide forest bird habitat quality score marginally increased while the grassland bird diversity score decreased. These findings make sense given land use shifted modestly toward forest and urban land use and out of agriculture and grasslands between 1992 and 2001.



	Scenarios			
Landscape Habitat Quality Scores	**1992**	**2001**	**Change**	**% Change**
Forest Birds	1631292	1640813	9521	0.6
Grassland Birds	1054823	1041964	-12859	-1.2

Figure 2. Results from InVEST habitat quality model run in Minnesota for breeding forest and grassland birds. Left to right, the first map represents LULC in 1992, the second map represents pixel-level habitat quality scores for forest birds in 1992, and the third map represents pixel-level habitat quality scores for grassland birds in 1992. The table summarizes changes in aggregate pixel-level habitat quality scores across the state between 1992 and 2001 given LULC change.

Example validation: a comparison of Tier 1 InVEST Habitat Quality with Minnesota's Gap Analysis Program

Here we provide some validation of the Tier 1 habitat quality model in Minnesota. As stated previously, the InVEST Tier 1 habitat quality score serves as a biodiversity proxy for a given conservation target. An implicit assumption is that, as the habitat quality score in a pixel increases then ecological conditions for species persistence increases. In order to test this assumption we compare changes in Tier 1 habitat quality scores to predicted species distributions maps provided by the Minnesota Gap Analysis Program (GAP; Drotts et al. 2007). GAP is a US Geological Survey program whose mission is to "keep common species common by identifying those species and plant communities that are not adequately represented on existing conservation lands" (USGS 2010). GAP maps the predicted distributions of all terrestrial vertebrate species that breed or use habitat in the state. All known, probable, and possible occurrence are used to define species range limits. The information required to create these maps is extensive, compared to the data needs of InVEST. In addition to the LULC maps created from detailed vegetation data, information about the distributions of each species, lists of all native species, specimen collection records, range maps, and documented habitat affinities for each species are collected in a GAP analysis. Consequently, GAP represents the most comprehensive set of predictive species occurrence models in terms of species considered and geographic coverage, and thus, provides an excellent validation dataset.

The study area for this comparison was the Little Cannon watershed of south-central Minnesota. First we calculated habitat quality scores in each pixel for forest bird species using the Tier 1 InVEST model. For this analysis we used the 2001 NLCD map as the LULC map. Next we parameterized the threat and sensitivity tables specifically for forest birds based on expert opinion and relevant literature (Table 5 and 6, respectively). From the Minnesota GAP data we used a map of predicted species distributions of forest birds to estimate species richness in each pixel. Finally to compare the habitat quality output map to the GAP richness map we used zonal statistics within ArcGIS to estimate the mean and standard deviation of number of species in each habitat quality score bin (e.g., 0.0 = 0.00 – 0.09, 0.1 = 0.10 – 0.19, … , 0.90 = 0.09 – 0.99). For example, in pixels with a habitat score between 0 and 0.09 we found the average species richness value and information on the range of richness values found (as summarized by the standard deviation statistic). This analysis was repeated for each pixel-level habitat score bin.

Table 5. Weights and effective distances for threat sources.

THREAT	MAX_DIST	WEIGHT	DECAY
ccrop	4.0	0.8	0
ldev	4.0	0.6	0
mdev	5.0	0.8	0
hdev	5.0	1.0	0
prd	3.0	0.8	0
srd	2.0	0.7	0

Note: CCROP = row crop; LDEV = low-density development; MDEV = medium-density development; HDEV = high-density development; PRD = primary road; SRD = secondary road.

Table 6. Sensitivity to threat sources and habitat suitability weights for each LULC type parameterized for breeding forest interior songbirds.

LULC	NAME	L_CCROP	L_LDEV	L_MDEV	L_HDEV	L_PRD	L_SRD	HABITAT
11	WATER	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21	DEVELOPED OPENSPACE	0.20	0.30	0.40	0.50	0.50	0.40	0.40
22	DEVELOPMENT LOW DENSITY	0.20	0.30	0.40	0.50	0.50	0.40	0.30
23	DEVELOPMENT MEDIUM DENSITY	0.20	0.30	0.40	0.50	0.50	0.40	0.20
24	DEVELOPMENT HIGH DENSITY	0.00	0.00	0.00	0.00	0.00	0.00	0.01
31	BARREN	0.00	0.00	0.00	0.00	0.00	0.00	0.01
41	DECIDUOUS FOREST	0.70	0.70	0.80	0.90	0.80	0.60	1.00
42	EVERGREEN FOREST	0.70	0.70	0.80	0.90	0.80	0.60	1.00
43	MIXED FOREST	0.70	0.70	0.80	0.90	0.80	0.60	1.00
52	SCRUB_SHRUB	0.50	0.30	0.50	0.70	0.70	0.50	0.50
71	GRASSLAND_HERBACEOUS	0.60	0.60	0.70	0.80	0.70	0.50	0.20
81	PASTURE_HAY	0.70	0.70	0.80	0.90	0.80	0.60	0.20
82	ROW CROPS	0.00	0.40	0.50	0.60	0.40	0.40	0.01
90	WOODY WETLANDS	0.80	0.60	0.70	0.90	0.80	0.60	1.00
95	HERBECEOUS WETLANDS	0.60	0.60	0.70	0.80	0.70	0.50	0.20

Note: CCROP = row crop; LDEV = low-density development; MDEV = medium-density development; HDEV = high-density development; PRD = primary road; SRD = secondary road.

Results from the analysis are presented in Figure 3, which illustrates overall concordance between MN GAP predicted species richness, and InVEST habitat quality pixel scores. In other words as habitat quality increases in an area more species are predicted to occur. Furthermore our results reflect the common curvilinear shape of species-areas curves (e.g., Rosenzweig 1995) whereby small improvements in marginal habitat quality scores (from 0.1 to 0.3) can yield large increases in the number of species (from 5 to 20 species, respectively). In general this exercise suggests the InVEST habitat quality model can serve as a proxy for species persistence on the landscape.

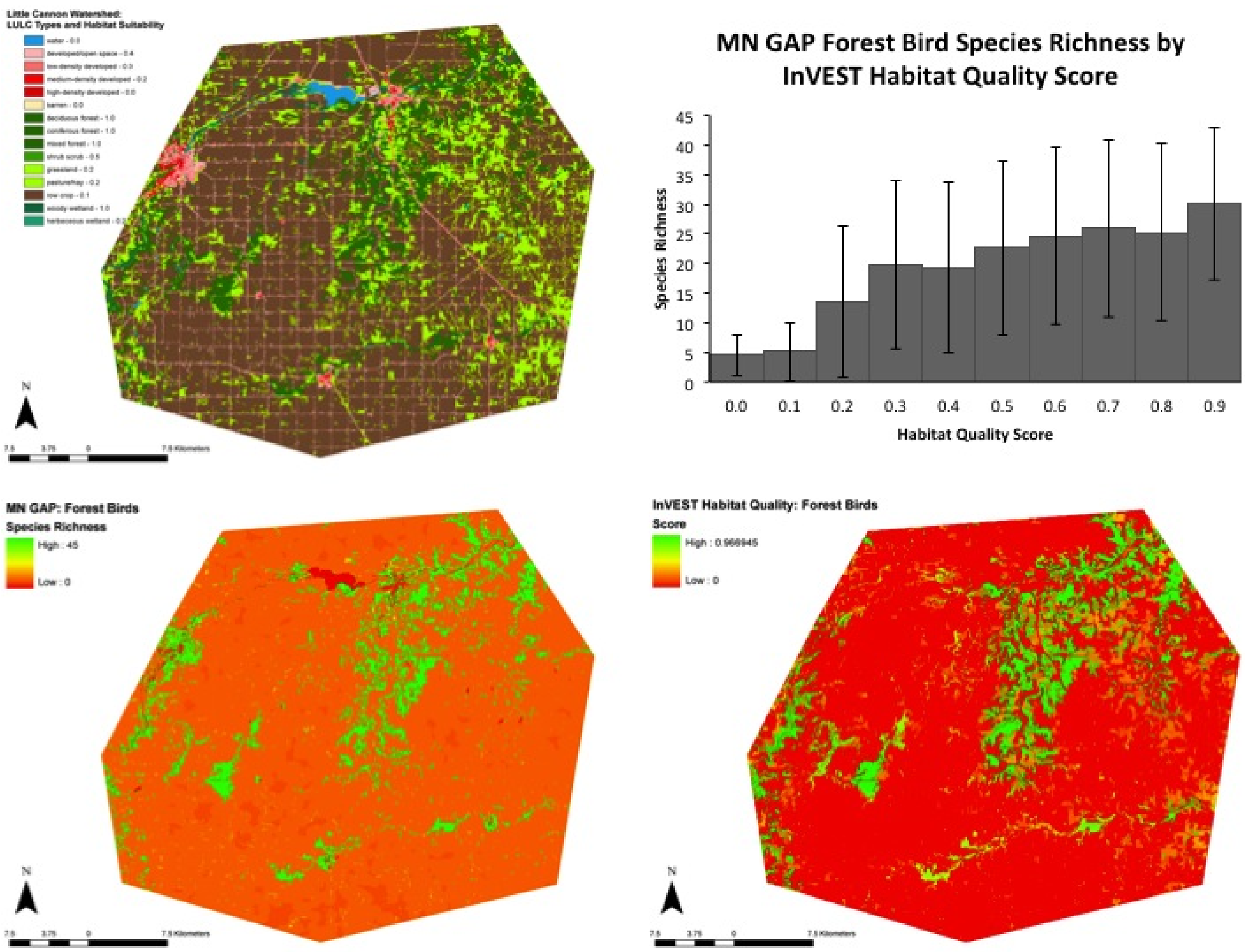


Figure 3. Results from a validation analysis of the InVEST habitat quality model based on forest bird diversity in Little Cannon watershed, Minnesota, U.S.A. Here the InVEST habitat quality output map was compared to the MN GAP predicted species richness map using zonal statistics within ArcGIS to estimate the mean and standard deviation of number of species for each habitat quality score bin (e.g., 0.0 = 0.00 – 0.09, 0.1 = 0.10 – 0.19, … , 0.90 = 0.09 – 0.99).

References

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

Audubon Minnesota. 2010. Minnesota watch list of birds at risk. Accessed February 5, 2011: http://mn.audubon.org/birds-science-education/state-birds/minnesota-watch-list-birds-at-risk.

Cunningham, M.A. and D.H. Johnson. 2006. Proximate and landscape factors influence grassland bird distributions. Ecological Applications 16:1062–1075.

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.

Ehrlich, P.R. D.S.W. Dobkin, and D. Wheye. 1988. The birder's handbook. a field guide to the natural history of North American birds. Simon and Schuster. New York, New York.

Donnelly, R. and J.M. Marzluff, 2004. Importance of reserve size and landscape context to urban bird conservation. Conservation Biology 18: 733–745.

Drotts, G., Heinzen, D. et al. 2007. The Minnesota gap analysis project final report. Division of Fish & Wildlife and Division of Forestry, Minnesota Department of Natural Resources, St.Paul, MN.

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

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

Fry, J.A., M.J. Coan, C.G. Homer, D.K. Meyer and J.D. Wickham. 2009. Completion of the National Land Cover Database (NLCD) 1992–2001 Land Cover Change Retrofit product. U.S. Geological Survey Open-File Report 2008–1379.

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 25:173-182.

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

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

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

Mayer, A.L. and G.N. Cameron. 2003. Landscape characteristics, spatial extent, and breeding bird diversity in Ohio, USA. Diversity and Distributions 9:297–311.

Minnesota Department of Natural Resources. 2000. Public land and mineral ownership in Minnesota: Minnesota Department of Natural Resources, Division of Lands and Minerals. 34 p.

Nelleman C, L. Kullered, ,.I. Vistnes , B. Forbes, T. Foresman, E. Husby, G. Kofinas, B. Kaltenborn, J. Rouaud, M. Magomedova, R. Bobiwash, C. Lambrechts, P. Schei, S. Tveitdal, O. Grøn, and T. Larsen. 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 and J. J. Lawler. 2008. Efficiency of incentives to jointly increase carbon sequestration and species

conservation on a landscape. Proceedings of the National Academy of Sciences 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. New York, New York.
Polasky, S., E. Nelson, D.N. Pennington and K. Johnson. 2011. 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 48:219-242.
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.
Rosenzweig, M. L. 1995 Species diversity in space and time. Cambridge University Press. New York, New York.
Sauer, J. R., J. E. Hines, and J. Fallon. 2008. The North American Breeding Bird Survey, Results and Analysis 1966 - 2007. Version 5.15.2008. USGS Patuxent Wildlife Research Center, Laurel, Maryland. http://www.mbr-pwrc.usgs.gov/bbs/.
USGS. 2010. What is Gap Analysis Program? Accessed February 5, 2010: http://www.nbii.gov/portal/server.pt/community/gap_home/1482.
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.

CARBON STORAGE AND SEQUESTRATION

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.

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 had prices 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

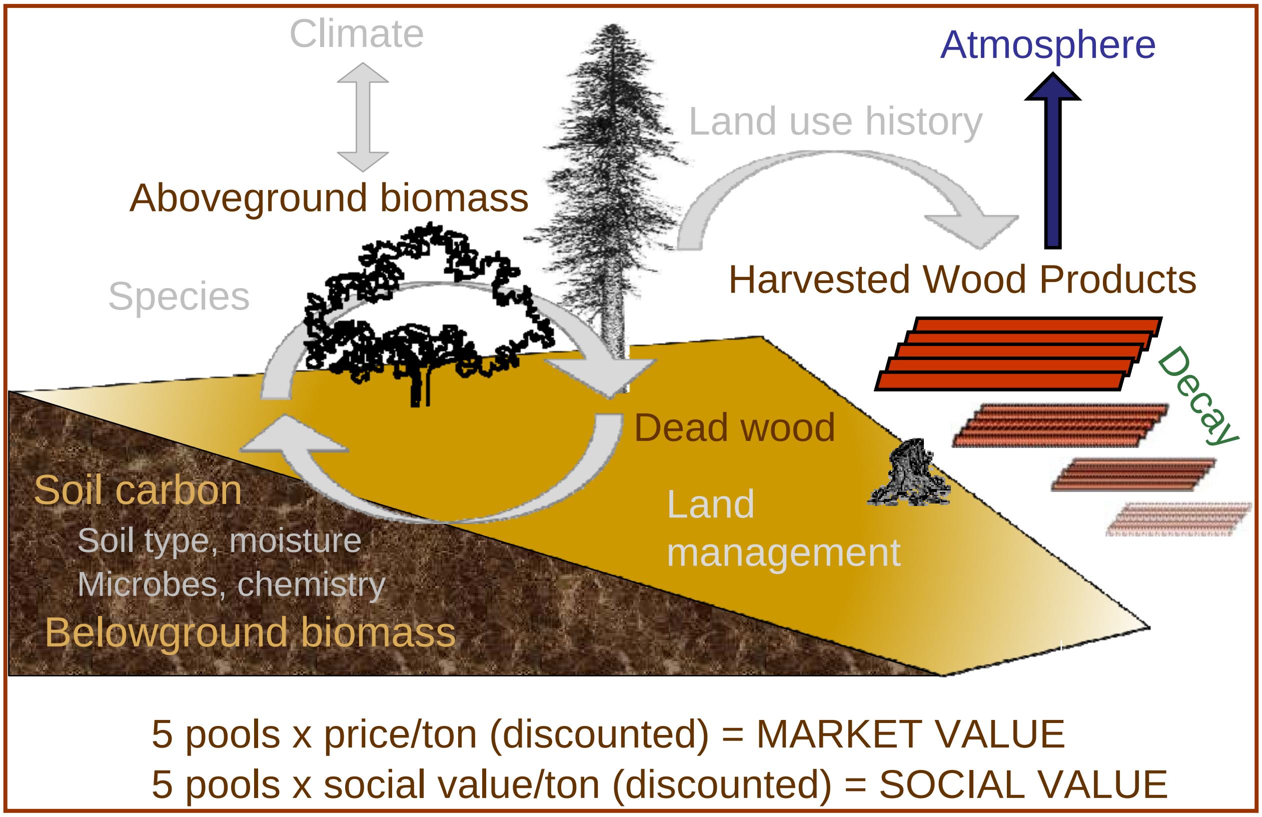


Figure 1. Conceptual model of carbon storage and sequestration. Parameters depicted in color are included in the InVEST model, while those in gray are not.

Birdsey 1994), in which case they could use the InVEST timber production model in tandem with the carbon model to assess management options.

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.

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.

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).

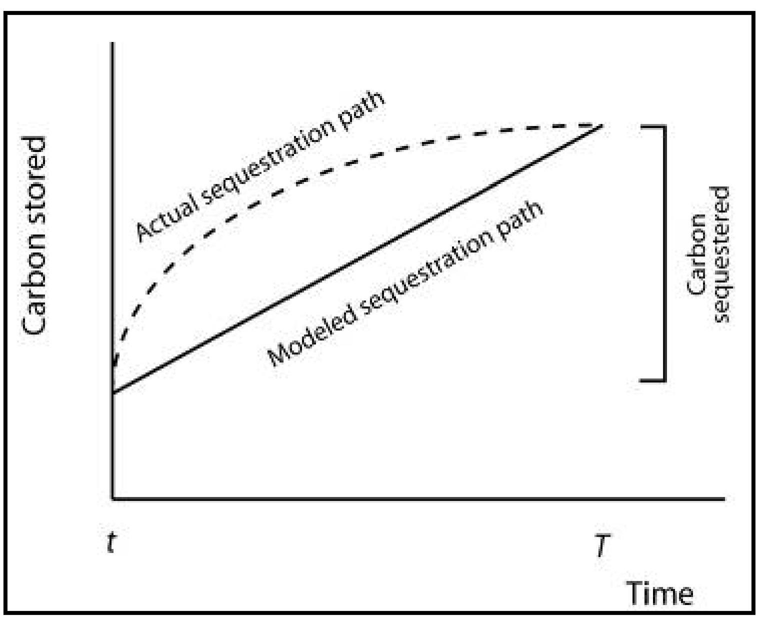


Figure 2*: The model assumes a linear change in carbon storage (the solid line), while the actual path to the year T’s carbon storage level may be non-linear (like the dotted line). In this case t can indicate the year of the current landscape and T the year of the future landscape. With positive discounting, 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.*

Data needs

The model uses five maps and tables of input data, two are required, and three are optional. 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**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 9/10^{ths} of the area occupied by plantation forests will be covered by trees at any point in time.

    Name: file can be named anything

    File type: *.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:

a. *LULC*: 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)
b. *LULC_name*: descriptive name of LULC class (optional)
c. *C_above*: amount of carbon stored in aboveground biomass (in Mg ha^{-1})
d. *C_below*: amount of carbon stored in belowground biomass (in Mg ha^{-1})
e. *C_soil*: amount of carbon stored in soil (in Mg ha^{-1})
f. *C_dead*: amount of carbon stored 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: \Invest\Carbon\Input**carbon_pools_samp.dbf**

Example: 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.

LULC	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

3. **Current harvest rates map (optional)**. A GIS shape file of polygons (parcels in our vernacular), contains data on:
   a. Parcel ID
   b. Amount of carbon, in the form of woody biomass, typically removed from the parcel over the course of a harvest period
   c. Date that the modeler wants to begin accounting for wood harvests in the parcel
   d. Frequency of harvest periods in the parcel in the past
   e. Average decay rate of products made from the wood harvested from a parcel
   f. Average carbon density of the wood removed form the parcel in the past
   g. 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:

- **a.** *FID*: unique identifying code for each polygon (parcels in our vernacular).
- **b.** *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).
- **c.** *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.
- **d.** *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.
- **e.** *Decay_cur*: The half-life of wood products harvested, measured in years.
- **f.** *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.
- **g.** *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 Mg of dry wood per m^3 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 $BCEF_R$ row in table 4.5 of IPCC (2006). Otherwise, set this expansion factor equal to 1 for each parcel.

Sample data set: \Invest\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 5^{th} 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 4^{th} 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:

$$HWP\_cur_x = Cut\_cur_x \times \sum_{t=0}^{\mathrm{ru}\left(\frac{yr\_cur - start\_date_x}{Freq\_cur_x}\right)-1} f\left(Decay\_cur_x; yr\_cur - start\_date_x - \left(t \times Freq\_cur_x\right)\right) \quad \text{(C1)}$$

where HWP_cur_x 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

$$f(\bullet) = \left\lfloor \frac{1 - e^{-\omega_x}}{\omega_x \times e^{\left[yr\_cur - start\_date_x - \left(t \times Freq\_cur_x\right)\right] \times \omega_x}} \right\rfloor, \quad \text{(C2)}$$

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

The following are several examples to show how equation (C1) works. In the first instance, assume $start\_date_x$ = 1983, *yr_cur* = 2000, and $Freq\_cur_x$ = 4. In this case, $\mathrm{ru}\left(\frac{yr\_cur - start\_date_x}{Freq\_cur_x}\right) = \mathrm{ru}\left(\frac{17}{4}\right) = \mathrm{ru}(4.25) = 5$. According to the summation term in equation (C1), 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 $yr\_cur - start\_date_x - \left(t \times Freq\_cur_x\right)$ to convert the series of *t*'s into years since harvest).

Alternatively, if $start\_date_x$ = 1980, *yr_cur* = 2000, and $Freq\_cur_x$ = 2 then $\mathrm{ru}\left(\frac{yr\_cur - start\_date_x}{Freq\_cur_x}\right) = \mathrm{ru}(10) = 10$. Therefore, according to equation (C1), harvests that contained Cut_cur_x 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_cur* for parcel *x* is measured in Mg (dry matter) ha^{-1} and is given by:

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

(C3)

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

$$Vol\_HWP\_cur_x = Bio\_HWP\_cur_x \times \frac{1}{Vol\_exp\_cur_x} \qquad \text{(C4)}$$

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.

   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 round down (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:*
   > (C:\InVEST\Base_Data**lulc_samp_fut)** and future harvest rate map
   > (C:\InVEST\Carbon**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 new *Cut* 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:

$$HWP\_fut_x = \left(Cut\_cur_x \times \sum_{t=0}^{\mathrm{ru}\left(\frac{\frac{yr\_fut+yr\_cur}{2}-start\_date_x}{Freq\_cur_x}\right)-1} f\left(Decay\_cur_x, yr\_fut - start\_date_x - \left(t \times Freq\_cur_x\right)\right)\right) + \left(Cut\_fut_x \times \sum_{t=0}^{\mathrm{ru}\left(\frac{yr\_fut-\frac{yr\_fut+yr\_cur}{2}}{Freq\_fut_x}\right)-1} f\left(Decay\_fut_x, yr\_fut - \frac{yr\_fut + yr\_cur}{2} - \left(t \times Freq\_fut_x\right)\right)\right) \quad \text{(C5)}$$

where the function *f* is as before. Recall that if (*yr_cur* + *yr_fut*) / 2 results in a fraction it is rounded down. Also note that equation (C5) does *not* include a harvest that is scheduled to occur in the future year; this harvest's carbon is *in 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 (C5):

$$HWP\_fut_x = Cut\_cur_x \times \sum_{t=0}^{\text{ru}\left(\frac{\frac{yr\_fut+yr\_cur}{2}-start\_date_x}{Freq\_cur_x}\right)-1} f\left(Decay\_cur_x, yr\_fut - start\_date_x - \left(t \times Freq\_cur_x\right)\right) \tag{C6}$$

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*:

$$HWP\_fut_x = Cut\_fut_x \times \sum_{t=0}^{\text{ru}\left(\frac{yr\_fut-\frac{yr\_fut+yr\_cur}{2}}{Freq\_fut_x}\right)-1} f\left(Decay\_fut_x, yr\_fut - \frac{yr\_fut+yr\_cur}{2} - \left(t \times Freq\_fut_x\right)\right) \tag{C7}$$

Note that this is the second term of equation (C5).

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:

$$Bio\_HWP\_fut_x = \left(Cut\_cur_x \times \text{ru}\left(\frac{\frac{yr\_fut+yr\_cur}{2}-start\_date_x}{Freq\_cur_x}\right) \times \frac{1}{C\_den\_cur_x}\right) + \left(Cut\_fut_x \times \text{ru}\left(\frac{yr\_fut-\frac{yr\_fut+yr\_cur}{2}}{Freq\_fut_x}\right) \times \frac{1}{C\_den\_fut_x}\right) \tag{C8}$$

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 (C8):

$$Bio\_HWP\_fut_x = \left(Cut\_cur_x \times \text{ru}\left(\frac{\frac{yr\_fut+yr\_cur}{2}-start\_date_x}{Freq\_cur_x}\right) \times \frac{1}{C\_den\_cur_x}\right) \tag{C9}$$

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* to *yr_fut* is given by second term of equation (C8):

$$Bio\_HWP\_fut_x = \left(Cut\_fut_x \times \text{ru}\left(\frac{yr\_fut-\frac{yr\_fut+yr\_cur}{2}}{Freq\_fut_x}\right) \times \frac{1}{C\_den\_fut_x}\right) \tag{C10}$$

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

$$Vol\_HWP\_fut_x = \left(Cut\_cur_x \times \mathrm{ru}\left(\frac{\frac{yr\_fut + yr\_cur}{2} - start\_date_x}{Freq\_cur_x} \right) \times \frac{1}{C\_den\_cur_x} \times \frac{1}{BCEF\_cur_x} \right) + \qquad , \qquad (C11)$$

$$\left(Cut\_fut_x \times \mathrm{ru}\left(\frac{yr\_fut - \frac{yr\_fut + yr\_cur}{2}}{Freq\_fut_x} \right) \times \frac{1}{C\_den\_fut_x} \times \frac{1}{BCEF\_cur_x} \right)$$

$$Vol\_HWP\_fut_x = \left(Cut\_cur_x \times \mathrm{ru}\left(\frac{\frac{yr\_fut + yr\_cur}{2} - start\_date_x}{Freq\_cur_x} \right) \times \frac{1}{C\_den\_cur_x} \times \frac{1}{BCEF\_cur_x} \right), \qquad (C12)$$

or

$$Vol\_HWP\_fut_x = \left(Cut\_fut_x \times \mathrm{ru}\left(\frac{yr\_fut - \frac{yr\_fut + yr\_cur}{2}}{Freq\_fut_x} \right) \times \frac{1}{C\_den\_fut_x} \times \frac{1}{BCEF\_cur_x} \right), \qquad (C13)$$

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. **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, setting this 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** is given by:

$$value\_seq_x = V \frac{sequest_x}{yr\_fut - yr\_cur} \sum_{t=0}^{yr\_fut - yr\_cur - 1} \frac{1}{\left(1+\frac{r}{100}\right)^t \left(1+\frac{c}{100}\right)^t} \qquad \text{(C14)}$$

Running the Model

Before running the Carbon Storage and Sequestration model, make sure that the INVEST toolbox has been added to your ARCMAP document, as described in the **Getting Started** chapter. Second, make sure that you have prepared the required input data files according to the specifications in Data Needs. Specifically, you will need (1) a land cover raster file showing the location of different land cover and land use types in the landscape; and (2) a carbon pools file which denotes the amount of aboveground, belowground, and soil carbon, and carbon from dead biomass, by land cover type. Optionally, you may also include (1) a map of harvest rates; (2) economic data on the value of carbon; and (3) future land use/land cover and harvest rate data to project future carbon scenarios.

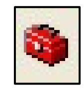
- Identify workspace

  If you are using your own data, you need to first create a workspace, or folder for the analysis data, on your computer hard-drive. The entire pathname to the workspace should not have any spaces. All your output files will be dumped here. For simplicity, you may wish to call the folder for your workspace "carbon" and create a folder in your workspace called “input” and place all your input files here. It’s not necessary to place input files in the workspace, but advisable so you can easily see the data you use to run your model.

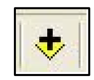
Or, if this is your first time using the tool and you wish to use sample data, you can use the data provided in InVEST-Setup.exe. If you unzipped the InVEST files to your C-drive (as described in the **Getting Started** chapter), you should see a folder called /Invest/**carbon**. This folder will be your workspace. The input files are in a folder called /Invest/carbon/**input** and in /Invest/**base_data**.

- Open an ArcMap document to run your model.

- Find the InVEST toolbox in ArcToolbox. ArcToolbox is normally open in ArcMap, but if it is not, click on the ArcToolbox symbol. See the **Getting Started** chapter if you don’t see the InVEST toolbox and need instructions on how to add it.

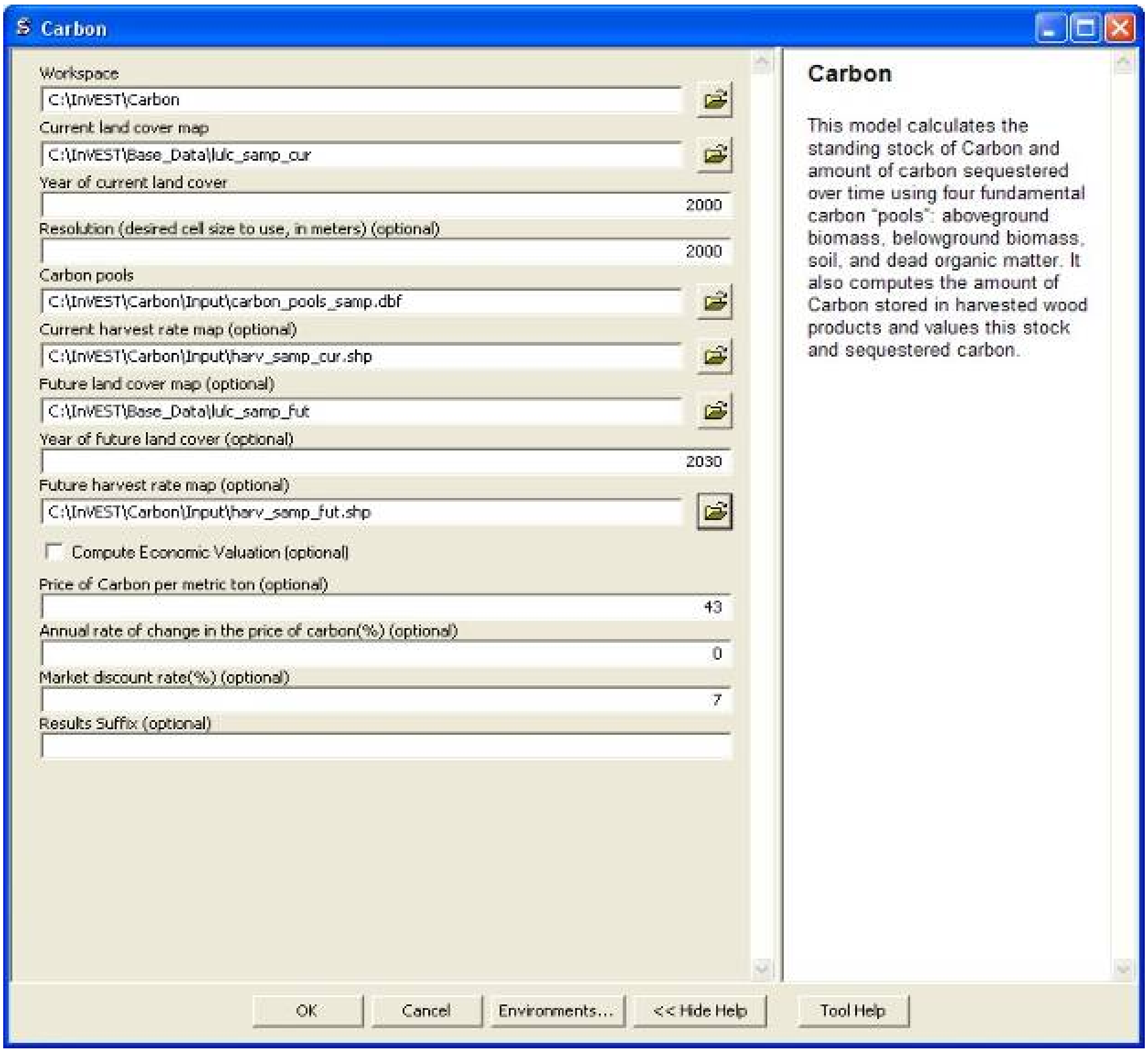


- You can run this analysis without adding data to your map view, but usually it is recommended to view your data first and familiarize yourself. Add the data for this analysis to your map using the Add Data button and look at each file to make sure it is formatted correctly. Save your ArcMap file as needed.



- Click once on the + sign on the left side of the InVEST toolbox to expand the list of tools. Double-click on Carbon.

Carbon tool dialog



- An interface will pop up like the one above. The tool shows default file names, but you can use the file buttons to browse instead to your own data. When you place your cursor in each space, you can read a description of the data requirements in the right side of the interface. In addition, refer to the *Data Needs* section above for information on data formats.

- Fill in data file names and values for all required prompts. Unless the space is indicated as optional, it requires you to enter some data. If you choose to run the optional economic valuation, all optional inputs below the checkbox become required.

- After you've entered all values as required, click on OK. The script will run, and its progress will be indicated by a "Progress dialogue".

- Upon successful completion of the model, you will see new folders in your workspace called "intermediate" and "output." These folders contain several raster grids. These grids are described in the *Interpreting Results* section.

- Load the output grids into 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 file appears in the map.

- You can also view the attribute data of output files by right clicking on a layer and selecting OPEN ATTRIBUTE TABLE.

Interpreting Results

Parameter log

Each time the model is run, a text 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.

Final results

Final results are found in the "Output" folder within the working directory set up for this model.

- ***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.

- ***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.

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.

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).

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/vol4.html, 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 estimates for 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://ncasi.uml.edu/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, $VOB \times WD \times BEF \times CF$ 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 ($WD \times BEF$) 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 $VOB \times WD \times BEF \times CF$ 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., $VOB \times WD \times BEF$). 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 ***not*** *treat 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 <u>savanna</u> 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 grassland LULC 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.ngdc.noaa.gov/seg/cdroms/ged_iia/datasets/a06/lh.htm). Silver *et al*. (2000) have estimated soil carbon as 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 aboveground 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 removed from 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 message boards at http://invest.ecoinformatics.org

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. The Asian Development Bank uses a rate of 10% to 12% when evaluating projects (http://www.adb.org/Documents/Guidelines/Eco_Analysis/discount_rate.asp). 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).

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://www.med.govt.nz/templates/MultipageDocumentTOC_28641.aspx>.

Adams, DM, RJ Alig, BA McCarl, *et al*. 1999. Minimum cost strategies for sequestering carbon in forests. *Land Econ* **75**: 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 Economics* **85**: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 Journal* **66**: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 Pollution* **116**: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 Applications* **9**: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. *Oecologia* **111**:1-11.

Canadell, JG and MR Raupach. 2008. Managing Forests for Climate Change Mitigation. *Science* **320**: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 Management* **164**:89-108.

Conte, MN and MJ Kotchen. 2010. Explaining the price of voluntary carbons offsets. *Climate Change Economics* (forthcoming).

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. *Biotropica* **30**:2-11.

Detwiler, RP. 1986. Land Use Change and the Global Carbon Cycle: The Role of Tropical Soils. *Biogeochemistry* **2**: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. *Science* **319**:1235-1238.

Feng, H. 2005. The dynamics of carbon sequestration and alternative carbon accounting, with an application to the upper Mississippi River Basin. *Ecological Economics* **54**: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 Biology* **4**:97-114.

Glenday, J. 2006. Carbon storage and emissions offset potential in an East African tropical rainforest. *Forest Ecology and Management* **235**:72-83.

Grace, J., J. San Jose, P. Meir, HS Miranda, and RA Montes. 2006. Productivity and carbon fluxes of tropical savannas. *Journal of Biogeography* **33**: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 Letters* **2**: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 Research* **111**.

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. *Science* **304**: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 Biology* **12**: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 Science* **7**:230-242.

McLauchlan, KK., SE Hobbie, and WM Post. 2006. Conversion From Agriculture To Grassland Builds Soil Organic Matter On Decadal Timescales. *Ecological Applications* **16**: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 Change* **83**: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 Science* **16**:78-93.

Murphy, JM *et 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 Change* **80**:127-143.

Nascimento, HEM, and WF Laurance. 2002. Total aboveground biomass in central Amazonian rainforests: a landscape-scale study. *Forest Ecology and Management* **168**: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 Environment* **Forthcoming**.

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. *Nature* **298**:156-159.

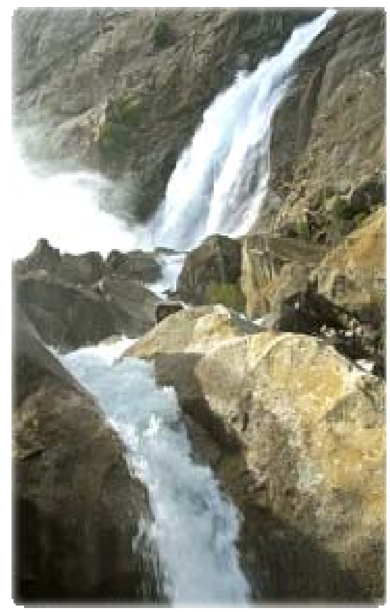
Post, WM, KC Kwon. 2000. Soil carbon sequestration and land-use change: processes and potential. *Global Change Biology* **6**:317-327.
Raich, JW, AE Russell, K. Kitayama, WJ Parton, and PM Vitousek. 2006. Temperature influences carbon accumulation in moist tropical forests. *Ecology* **87**:76-87.
Ruesch 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 Ecology* **8**: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 Change* **40**:105-117.
Tilman, D., J. Hill, and C. Lehman. 2006. Carbon-Negative Biofuels from Low-Input High-Diversity Grassland Biomass. *Science* **314**:1598-1600.
Tol, RSJ. 2005. The marginal damage costs of carbon dioxide emissions: an assessment of the uncertainties. *Energy Policy* **33**:2064-2074.
Tol, RSJ. 2009. The Economic Effects of Climate Change. *Journal of Economic Perspectives* **23**: 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 & Development* **16**:53-71.
Weitzman, ML. 2007. A review of the Stern Review on the Economics of Climate Change. *Journal of Economic Literature* **45**:703-724.

Zhang, Q, and CO Justice. 2001. Carbon Emissions and Sequestration Potential of Central African Ecosystems. *AMBIO* **30**:351-355.

RESERVOIR HYDROPOWER PRODUCTION

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, soil 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 mode assumes that energy pricing is static over time.



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.

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 (e.g. WEAP model). Sophisticated models of these connections and associated processes are resource and data intensive and require substantial expertise. To accommodate more contexts, for which data is readily available data, 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 sub-basin 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 sub-basin 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 sub-watershed to watershed scale. We are only confident in the interpretation of these models at the sub-watershed scale, so all outputs are summed and/or averaged to the sub-basin 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.**

Beyond annual average runoff, we also calculate 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_{jx}) for each pixel on the landscape (indexed by $x = 1,2,\ldots,X$) as follows:

$$Y_{xj} = \left(1 - \frac{AET_{xj}}{P_x}\right) \cdot P_x$$

where, AET_{xj} is the annual actual evapotranspiration on pixel x with LULC j and P_x is the annual precipitation on pixel x.

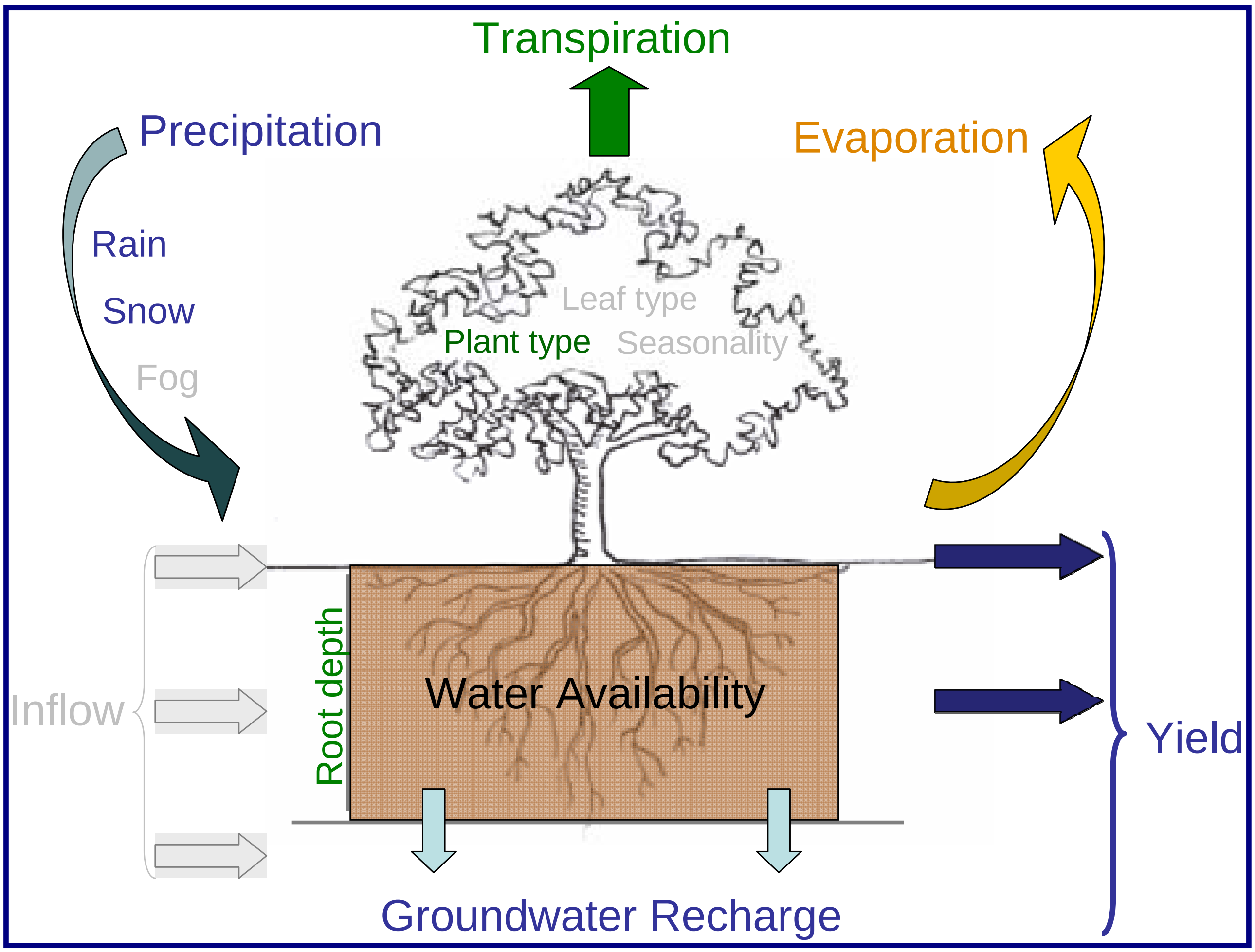


Figure 1. Conceptual diagram of the water balance model used in the hydropower production model. The water cycle is simplified, including only the parameters shown in color, and ignoring the parameters shown in gray. Yield, as calculated by this step of the model, is then adjusted for other consumptive uses and applied to hydropower energy and value estimates.

The evapotranspiration partition of the water balance, $\frac{AET_{xj}}{P_x}$, is an approximation of the Budyko curve developed by Zhang *et al*. (2001):

$$\frac{AET_{xj}}{P_x} = \frac{1+\omega_x R_{xj}}{1+\omega_x R_{xj} + \frac{1}{R_{xj}}}$$

where, R_{xj} is the dimensionless Budyko Dryness index on pixel *x* with LULC *j*, defined as the ratio of potential evapotranspiration to precipitation (Budyko 1974) and ω_x is a modified dimensionless ratio of plant accessible water storage to expected precipitation during the year. As defined by Zhang *et al*. (2001) ω_x is a non-physical parameter to characterize the natural climatic-soil properties.

$$\omega_x = Z\frac{AWC_x}{P_x}$$

where AWC_x is the volumetric (mm) plant available water content. The soil texture and effective soil depth define AWC_x, which establishes the amount of water that can be held and released in the soil for use by a plant, estimated as the product of the difference between field capacity and wilting point and the minimum of soil depth and root depth. *Z* is a seasonality factor that presents the seasonal rainfall distribution and rainfall depths. In areas of winter (December to April) rains, we expect to have Z on the order of 10, in humid areas with rain events distributed throughout the year or regions with summer rains the Z is on the order of 1. While we calculate ω_x, in some cases specific biome values already exist based on water availability and soil-water storage (Milly 1994, Potter et al. 2005, Donohue et al. 2007).

Finally, we define the Budyko dryness index, where R_{xj} values that are greater than one denote pixels that are potentially arid (Budyko 1974), as follows:

$$R_{xj} = \frac{k_{xj} \cdot ETo_x}{P_x}$$

where, ETo_x is the reference evapotranspiration from pixel *x* and k_{xj} is the plant (vegetation) evapotranspiration coefficient associated with the LULC *j* on pixel *x*. ETo_x represents an index of climatic demand while k_{xj} is largely determined by *x*'s vegetative characteristics (Allen *et al*. 1998).

The water yield model script generates and outputs the total and average water yield at the sub-basin level.

Water Scarcity Model

The Water Scarcity Model calculates the water scarcity value along a flow path based on water yield and water consumptive use in a watershed(s) of interest. The user inputs how much water is consumed by each land use land cover type in a table format. 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 water used for cooling or other industrial processes that return water to the stream after use. 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 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*.

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}. Divide the observed value by the estimated value to derive a calibration constant. This can then be entered in to the hydropower station data table and used to make power and value estimates actual rather than relative.

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 or the revenue is then redistributed over the landscape based on the proportional contribution of each parcel to energy production. Final output maps show how much water yield from each pixel of the landscape contributes to hydropower production, and how much energy production or hydropower value can be attributed to that water yield over the lifetime of the reservoir.

At dam *d*, power is calculated using the following equation:

$$p_d = \rho . q_d . g . h_d$$

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:

$$\varepsilon_d = 0.00272.\beta.\gamma_d.h_d.Vin$$

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,

$$NPVH_d = \left(p_e\varepsilon_d - TC_d\right) \times \sum_{t=0}^{T-1} \frac{1}{(1+r)^t}$$

where TC_d is the total annual operating costs for dam *d*, p_e is the market value of electricity (per unit of energy consumed) 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 , ε_d , and are constant over time.

Once the total water yield contributing to annual energy production, the average amount of energy produced by that inflow volume each year, and the net present value of that energy production are determined, we need to allocate these results back to each pixel on the landscape, in proportion to their contribution to water yield.

First, we determine each pixel's contribution to the total water yield used for hydropower production as:

$$c_x = \frac{Y_{jx} - u_x}{V_{in}}$$

where Y_{jx} is the water yield at each pixel, u_x is the consumptive use of the pixel x, and V_{in} the inflow water into dam *d.* This output represents the ecosystem service in biophysical terms: provision of water for hydropower production. The following equations translate this level of ecosystem service provision into energy units and dollar value units.

Energy production over the lifetime of dam *d* is attributed to each pixel as follows:

$$\varepsilon_x = \left(T_d\varepsilon_d\right) \times c_x$$

where the first term in parentheses represents the electricity production over the lifetime of dam *d*. The second term represents the proportion of water used for hydropower production that comes from pixel *x*.

The value of each pixel for hydropower production over the lifetime of dam d is calculated similarly:

$$NPVH_x = NPVH_d \times c_x$$

The energy production and the energy value are then summed to the sub-basin level.

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 assumes that all water produced in a watershed in excess of evapotranspiration arrives at the watershed outlet, without considering water capture by means other than primary human consumptive uses. Surface water – ground water interactions are entirely neglected, which may be a cause for error especially in areas of karst geology. The relative contribution of yield from various parts of the watershed should still valid.

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 more likely to affect the timing of flows than the annual water yield, and are more of a 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 describes consumptive demand by LULC type. 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 much further upstream in a rural location. Spatial disparity in actual and modeled demand points may cause an incorrect representation in the scarcity 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, 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.

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.

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. Soil depth (required). A GIS raster dataset with an average soil depth value for each cell. The soil depth values should be in millimeters.

Name: File can be named anything, but no spaces in the name and less than 13 characters
Format: Standard GIS raster file, with an average soil depth in millimeters for each cell.
Sample data set: \InVEST\ Base_Data **soil_depth**

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
Format: Standard GIS raster file (e.g., ESRI GRID or IMG), with precipitation values for each cell.
Sample data set: \InVEST\ Base_Data **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
Format: Standard GIS raster file (e.g., ESRI GRID or IMG), with available water content values for each cell.
Sample data set: \InVEST\Base_Data**pawc**

4. Average Annual Potential Evapotranspiration (required). A GIS raster dataset, with an annual average evapotranspiration value for each cell Potential 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 evapotranspiration values should be in millimeters.

Name: File can be named anything, but no spaces in the name and less than 13 characters
Format: Standard GIS raster file (e.g., ESRI GRID or IMG), with potential evapotranspiration values for each cell.

Sample data set: \InVEST\Base_Data**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
Format: Standard GIS raster file (e.g., ESRI GRID 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**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**watersheds.shp**

7. **Sub-watersheds (required).** A shapefile, with one polygon per sub-watershed within the main watersheds specified in the Watersheds shapefile. See the Working with the DEM section for information about generating sub-watersheds.

Format: Shapefile (.shp)
Rows: Each row is one sub-watershed
Columns: An integer field named *subws_id* is required, with a unique integer value for each sub-watershed
Sample data set: \InVEST\ Base_Data\sub**watersheds.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\Water_Tables.mdb**Biophysical_Models**

Name: Table names should only have letters, numbers and underscores, no spaces
Format: .dbf or .mdb
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:

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 above.

b. *LULC_desc*: Descriptive name of land use/land cover class (optional)

c. *root_depth:* The maximum root depth for vegetated land use classes, given in integer millimeters. Non-vegetated LULCs should be given a value of 0.

d. *etk:* The plant evapotranspiration coefficient for each LULC class, used to obtain potential evapotranspiration by using plant energy/transpiration characteristics to modify the reference evapotranspiration, which is based on alfalfa. Coefficients should be multiplied by 1000, so that the final *etk* values given in the table are integers ranging between 1 and 1500 (some crops evapotranspire more than alfalfa in some very wet tropical regions and where water is always available).

9. Zhang constant (required): Floating point value between 1 and 10 corresponding to the seasonal distribution of precipitation (see Appendix A).

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\Water_Tables.mdb**Water_Demand**

Name: Table names should only have letters, numbers and underscores, no spaces

Format: .dbf or .mdb

Rows: Each row is a landuse / landcover class

Columns: Contain water demand values per LULC class and must be named as follows:

a. *lucode*: Integer value of land use/land cover class (e.g., 1 for forest, 3 for grassland, etc.), must match LULC raster, described above.

b. *demand*: The estimated average consumptive water use for each landuse / landcover type. Water use should be given in integer cubic meters per year.

11. Hydropower Table. A table of hydropower stations with associated model values.

Sample data set: \InVEST\Base_Data\Water_Tables.mdb**Hydropower_Valuation**

Name: Table names should only have letters, numbers and underscores, no spaces

Format: .dbf or .mdb

Rows: Each row is a hydropower station

Columns: Each column contains an attribute of each hydropower station, and must be named as follows:

a. *ws_id*: Unique integer value for each watershed, which must correspond to values in the Watersheds layer.

b. *station_desc*: Name of hydropower station (optional)

c. *calib*: Annual water yield calibration constant. Multiplying this value by the total water supply for a watershed should give the actual total annual water supply observed/measured at the point of interest, corresponding to the 'wsupply' column of the Scarcity tool's 'water_scarcity.dbf' output. Floating point value.

d. *efficiency*: The turbine efficiency. A number to be obtained from the hydropower plant manager (floating point values generally 0.7 to 0.9)
e. *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, environmental demands. Floating point value.
f. *height*: The head, measured as the average annual effective height of water behind each dam at the turbine intake in meters. Floating point value.
g. *kw_price*: The price of one kilowatt-hour of power produced by the station, in dollars or other currency. Floating point value.
d. *cost*: Annual cost of running the hydropower station (maintenance and operations costs). Floating point value.
e. *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.
f. *discount:* The discount rate over the time span, used in net present value calculations. Floating point value.

12. Hydropower calibration table. A table of hydropower stations with associated calibration values.
Sample data set: \InVEST\Base_Data\Water_Tables.mdb**Hydropower_Calibration**

Name: Table names should only have letters, numbers and underscores, no spaces
Format: .dbf or .mdb
Rows: Each row is a hydropower station
Columns: Each column contains an attribute of each hydropower station, and must be named as follows:

a. *ws_id*: Unique integer value for each watershed, which must correspond to values in the Watersheds layer.
b. *calib*: Annual water yield calibration constant. Multiplying this value by the total water supply for a watershed should give the actual total annual water supply observed/measured at the point of interest, corresponding to the 'cyield' column of the Scarcity tool's 'water_scarcity.dbf' output. Floating point value.

Running the Model

The Hydropower model maps the water yield, water consumption, energy produced by water yield and corresponding energy value over the landscape. This model is structured as a toolkit which has three tools. The first tool, **Water Yield**, calculates the surface water yield and actual evapotranspiration across the landscape. This output feeds into the next portion of the model, the **Water Scarcity** tool, which calculates water consumption, supply and realized supply, which is yield minus consumption. The third tool, **Valuation**, calculates energy production and the value of that energy, as it can be attributed to sub-basins on the watershed of interest.

By running the tool, three folders will automatically be created in your workspace (you will have the opportunity to define this file path), "Intermediate", where temporary files are written, and which is deleted after each tool run; "Service", where results that show ecosystem services are saved; and "Output", where non-service biophysical results are saved.

Before running the Hydropower Model, make sure that the InVEST toolbox has been added to your ArcMap document, as described in the **Getting Started** chapter of this manual. Second, make sure that you have prepared the required input data files according to the specifications in *Data Needs*.

- Identify workspace

  If you are using your own data, you need to first create a workspace, or folder for the analysis data, on your computer hard-drive. The entire pathname to the workspace should not have any spaces. All your output files will be saved here. For simplicity, you may wish to call the folder for your workspace 'Hydropower' and create a folder in your workspace called "Input" and place all your input files here. It's not necessary to place input files in the workspace, but advisable so you can easily see the data you use to run your model.

  Or, if this is your first time using the tool and you wish to use sample data, you can use the data provided in InVEST-Setup.exe. If you unzipped the InVEST files to your C drive (as described in the **Getting Started** chapter), you should see a folder called /Invest/Hydropower. This folder will be your workspace. The input files are in a folder called /Invest/Base_Data/Freshwater.

- Open an ArcMap document to run your model.

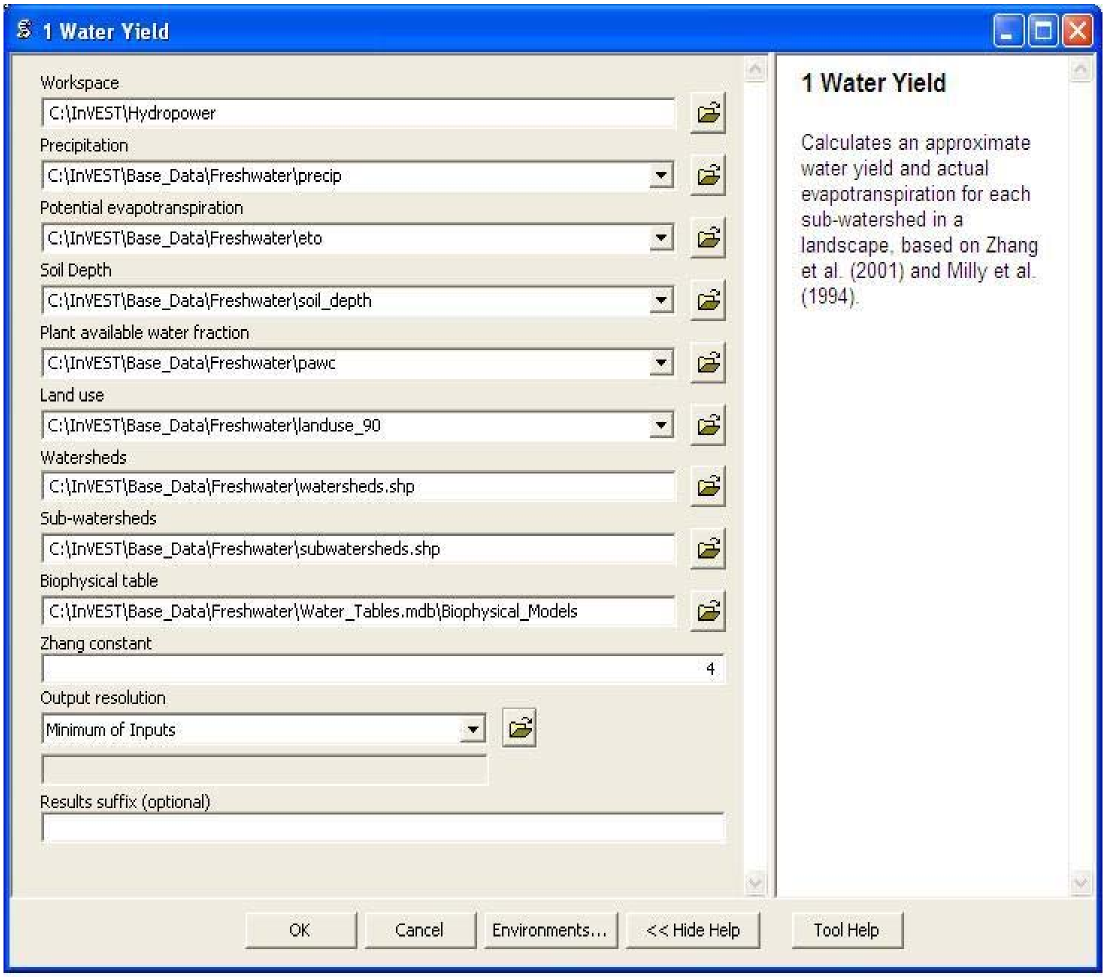
- Find the InVEST toolbox in ArcToolbox. ArcToolbox is normally open in ArcMap, but if it is not, click on the ArcToolbox symbol. See the **Getting Started** chapter if you don't see the InVEST toolbox and need instructions on how to add it.

- You can run this analysis without adding data to your map view, but usually it is recommended to view your data first and get to know them. Add the data for this analysis to your map using the ADD DATA button and look at each file to make sure it is formatted correctly. Save your ArcMap file as needed.

- Click once on the plus sign on the left side of the InVEST toolbox to see the list of tools expand. Next, click on the plus sign next to the **Hydropower** toolset. Within the toolset are three tools, **Water Yield**, **Water Scarcity** and **Valuation**. You will need to run **Water Yield** first, **Water Scarcity** second and **Valuation** last, as each tool generates outputs that feed into the next.

- Double click on **Water Yield**. An interface will pop up like the one below. The tool shows default file names, but you can use the file buttons to browse instead to your own data. When you place your cursor in each space, you can read a description of the data requirements in the right side of the interface. ***Click Show Help if the description is not***

displayed. In addition, refer to the *Data Needs* section above for information on data formats.



- Fill in data file names and values for all required prompts. Unless the space is indicated as optional, it requires you to enter some data.

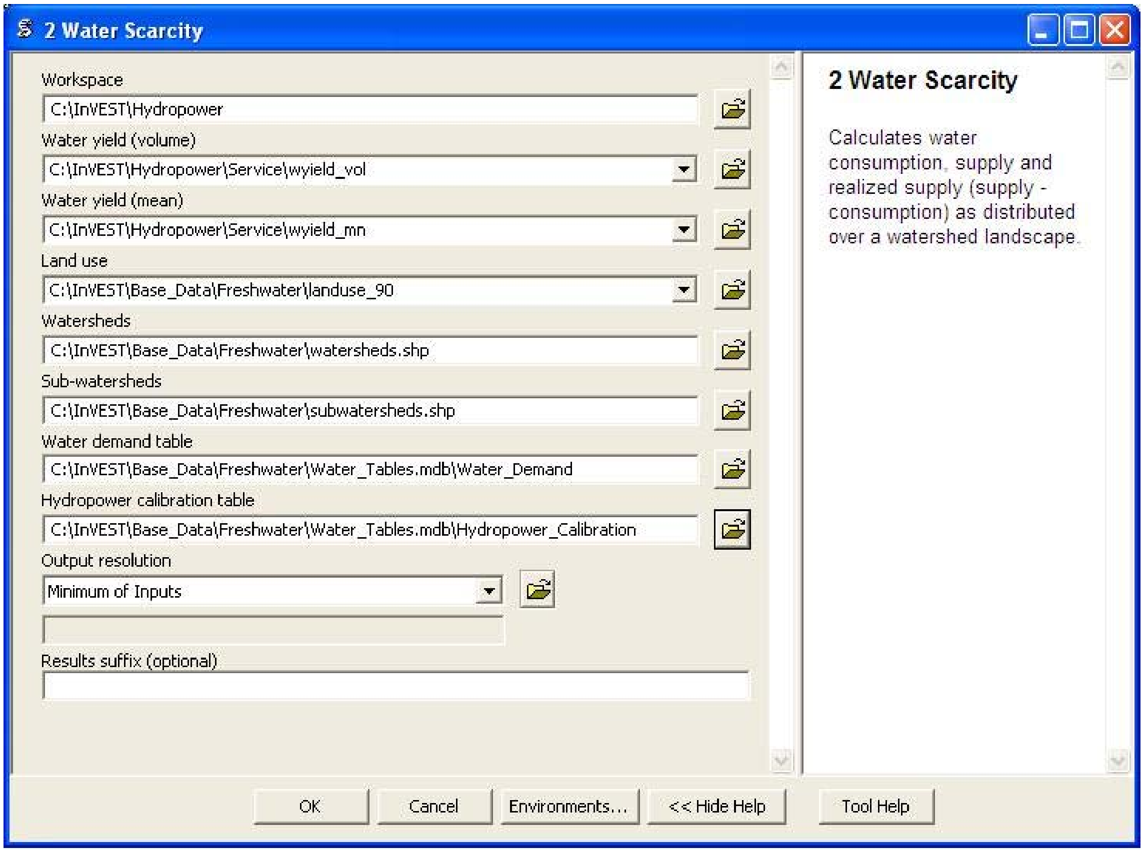
- After you've entered all values as required, click on OK. The script will run, and its progress will be indicated by a "Progress dialogue".

- Load the output grids into ArcMap using the ADD DATA button from either "Output" or "Service" folders.

- 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 file appears in the map. You may change the coloring scheme for better visualization.

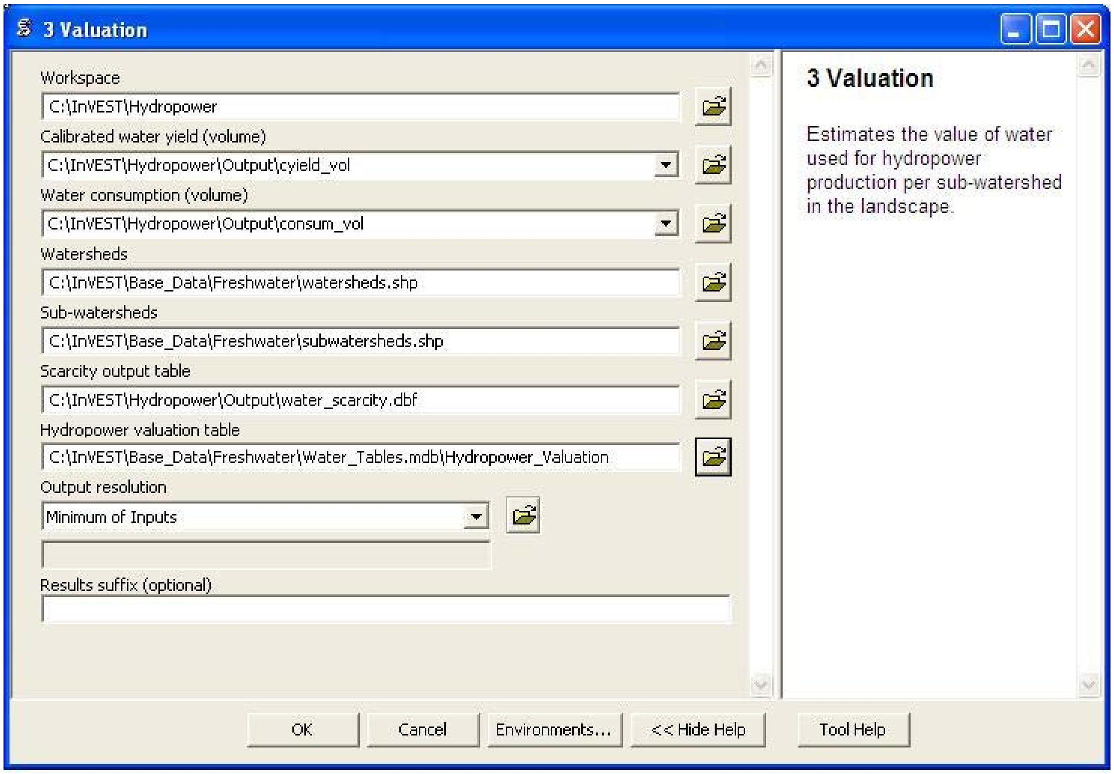
- You can also view the attribute data of output files by right clicking on a layer and selecting OPEN ATTRIBUTE TABLE.

- Now, run the tool **Water Scarcity**. Two outputs from the Water Yield model, **wyield_vol and wyield_mn**, serve as inputs to this model and are found in the service folder (see results interpretation section). The interface is below:



- When the script completes running, its results will be saved in the Output folder. A description of these results is in the next section. Load them into your ArcMap project, look at them, and check out the attribute table.

- Finally, run the tool **Valuation**. Three outputs from previous tools are required: **cyield_vol, consum_vol**, and **water_scarcity.dbf** from Scarcity. The interface is below:



- When the script completes running, its results will be saved in the Service folder. A description of these results is in the next section. Load them into your ArcMap project, look at them, and check out the attribute table.

This model is open source, so you can edit the scripts to modify, update, and/or change equations by right clicking on the script's name and selecting "Edit…" The script will open in a text editor. After making changes, click File/Save to save your new script.

Interpreting Results

The model runs on the pixel level then it sums and averages these outputs at the sub-basin level. In this section, we focus on describing the outputs at the sub-watershed level.

The following is a short description of each of the outputs from the Reservoir Hydropower Production tool (each of these output files is automatically saved in an "Output" or "Service" folder that is saved in the Working Directory that the user specifies):

1. Output\ **fractp_mn**: Mean actual evapotranspiration fraction of precipitation per sub-watershed (Actual Evapotranspiration / Precipitation). It is the mean fraction of precipitation that actually evapotranspires at the sub-basin level. Output**aet_mn**: Actual evapotranspiration per sub-watershed (in mm)
2. Service**wyield_vol**: Total water yield per sub-watershed. The approximate absolute annual water yield across the landscape, calculated as the difference between precipitation and actual evapotranspiration on each land parcel. Given in m^3
3. Service**wyield_mn**: Mean water yield per sub-watershed. Given in mm.

4. Service**wyield_ha**: Water yield volume per hectare per sub-watershed. Given in m^3/hectare.
5. Output**consum_vol**: Total water consumptive use for each sub-watershed, where each land use type is mapped to a corresponding value in the Water Demand table consump field. Given in m^3
6. Output**consum_mn**: Mean water consumptive volume per hectare per each sub-watershed. Given in m^3/hectare.
7. Output**rsupply_vol**: Realized water supply (water yield – consumption) volume for each sub-watershed (in m^3)
8. Output**rsupply_mn**: Mean realized water supply (water yield – consumption) per sub-watershed(in m^3/hectare).
9. Output**cyield_vol**: Calibrated water yield volume per sub-watershed (water yield * calibration constant) (in m^3).
10. Output\water_scarcity.dbf: Table containing values for total water demand, realized supply and calibrated yield for each watershed (in m^3)
11. Service**hp_en_ws**: **THIS IS THE WATERSHED MAP OF THIS ECOSYSTEM SERVICE IN ENERGY PRODUCTION TERMS**. This grid shows the amount of energy produced by the hydropower station over the specified time span that can be attributed to each watershed based on its water yield contribution.
12. Service**hp_en_sws**: **THIS IS THE SUB-WATERSHED MAP OF THIS ECOSYSTEM SERVICE IN ENERGY PRODUCTION TERMS**. This grid shows the amount of energy produced by the hydropower station over the specified time span that can be attributed to each sub-watershed based on its water yield contribution.
13. Service**hp_val_ws: THIS IS THE WATERSHED MAP 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 time span. (in the currency given in the Hydropower table.)
14. Service**hp_val_sws: THIS IS THE SUB-WATERSHED MAP OF THIS ECOSYSTEM SERVICE IN ECONOMIC TERMS.** This grid shows the value of the landscape per sub-watershed according to its ability to yield water for hydropower production over the specified time span. (in the currency given in the Hydropower table.)
15. Service**hydropower_value_subwatershed.dbf**: Table containing values for total water supply and demand for each sub-watershed, plus the total energy produced, and the value of the energy produced, per hydropower station. Values as specified above.

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 to evapotranspiration across the watershed.
The *wyield_vol* raster shows the annual average water volume that is 'yielded' from each sub-watershed of the watershed of interest. This raster can be used to determine which sub-watersheds 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) grid then shows how much water is used for consumptive activities (such as drinking, bottling, etc.) each year across the landscape. The realized supply (rsupply_vol) grid calculates the difference between cumulative water yield and cumulative consumptive use. This grid 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 grid 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 grid offers a general sense of the water balance and whether there is a lack of or abundance of water in the area of interest.

The hp_en_sws and hp_val_sws grids are the most relevant model outputs for prioritizing the landscape for investments that wish to maintain water yield for hydropower production. The hp_val_sws grid contains the most information for this purpose as it represents the revenue attributable to each sub-watershed over the expected lifetime of the hydropower station, or the number of years that the user has chosen to model. This grid 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 sub-watersheds watersheds contribute the highest value water for energy production. If energy values do not vary much across the landscape, the hp_en_sws outputs can be just as useful in planning and prioritization. Comparing any of these grids between landuse scenarios allows the user to understand how the role of the landscape may change under different management plans.

The hydropower output summary table presents the model results in terms of hydropower operation. The ‘wyield’ field provides the total volume of water that arrives at each hydropower plant every year, considering water yield and consumption. The 'consump' field provides the total volume of water that is consumed in each watershed upstream of the station. Total energy produced at each hydropower station is given in the 'energy' field, and the corresponding value of that energy is given in the 'value' field. This table provides a quick comparison between land use scenarios in a way that complements the spatial representation across the landscape. Ideally the output grids and summary table will be used together for comparison of land use and management scenarios.

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. 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 (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.

b. Average annual reference evapotranspiration (ET_o)

Reference evapotranspiration, ET_o, 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 (see http://mercury.ornl.gov/metadata/mastdc/html/nacp/daac.ornl.gov_data_bluangel_harvest_RGED_curtis_metadata_climate_monthly_evapotranspiration.html). Reference evapotranspiration depends on 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 (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'[5] equation, which generates superior results than the Pennman-Montieth when information is uncertain.

$$ETo = 0.0013 \times 0.408 \times RA \times (T_{av} + 17) \times (TD - 0.0123\ P)^{0.76}$$

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 MJm-2d-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[6] or equations[7].

The Potential evapotranspiration could be also calculated monthly and annually using the Hamon equation (Hamon 1961, Wolock and McCabe 1999):

$$PET_{Hamon} = 13.97 d D^2 W_t$$

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:

$$W_t = \frac{4.95 e^{0.062T}}{100}$$

where T is the monthly mean temperature in degrees Celsius. Potential evapotranspiration was set to zero when mean monthly temperature was below zero. Then for each year during the time periods analyzed, the monthly calculated PET values at each grid cell were summed to calculate a map of the annual PET for each year.

---

5 Droogers P., and R.G. Allen. 2002. Estimating reference evapotranspiration under inaccurate data conditions. *Irrigation and Drainage Systems* 16: 33-45.

6 Hargreaves G.H. 1994. Defining and using reference evapotranspiration. *J. Irrig. and Drain*. Engrg., ASCE 120(6): 1132–1139.

7 Allen R.G., Pereira L.S., Raes D. & Smith M. 1998. *Crop evapotranspiration: Guidelines for computing crop requirements. Irrigation and Drainage Paper* No. 56, FAO, Rome, Italy.

c. Soil depth

Soil depth may be obtained from standard soil maps. Coarse, yet free global soil characteristic data is available at http://www.ngdc.noaa.gov/seg/cdroms/reynolds/reynolds/reynolds.htm. The FAO also provides global soil data in their Harmonized World Soil Database: http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/
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://soils.usda.gov/survey/geography/ssurgo/ and STATSGO http://soils.usda.gov/survey/geography/statsgo/ . 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 depth should be calculated as the max depth of all horizons within a soil class component, and then a weighted average of the components should be estimated. This can be a tricky GIS analysis: In the US soil categories, each soil property polygon can contain a number of soil type components with unique properties, and each component may have different soil horizon layers, also with unique properties. Processing requires careful weighting across components and horizons. The Soil Data Viewer (http://soildataviewer.nrcs.usda.gov/), a free ArcMap extension from the NRCS, does this soil data processing for the user and should be used whenever possible. Ultimately, a grid layer must be produced. Data gaps, such as urban areas or water bodies need to be given appropriate values. Urban areas and water bodies can be thought of having zero soil depth. A good product would be to determine the minimum of depth to bedrock and typical water table depth.

d. Plant available water content (PAWC)

Plant available water content is a fraction obtained from most 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 'Soil 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.

e. 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.umiacs.umd.edu/data/landcover/. This data is available in 1 degree, 8km and 1km resolutions. 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 Landuse Land class Table.

ID	Land Use /Land Class
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

f. Maximum root depth table

A valuable review of maximum plant rooting depths is available in Canadell, J., R. B. Jackson, and H. Mooney. 1996, Maximum rooting depth of vegetation types at the global scale. *Oecologia* 108: 583-595 where 290 observations in the literature are summarized, and it is concluded that rooting depths are more consistent than that previously believed among similar biomes and plant species. The model determines the minimum of soil depth and rooting depth for an accessible soil profile for water storage. Determinations on how to deal with soil-less systems, such as fractured rock substrates, should be based on expert advice. Effective maximum root depth must be defined for impermeable landuse/land classes, such as urban areas, or water bodies. A rule of thumb is to denote water and urban areas with minimal maximum rooting depths, but a zero value should not be used. The literature values must be converted to mm, and depicted as integer values.

Maximum root depths by species and biomes

Root Depth by Species	Root Depth by Biome
Trees 7.0 m	Cropland 2.1 m
Shrubs 5.1 m	Desert 9.5 m
Herbaceous Plants 2.6 m	Sclerophyllous Shrubland & Forest 5.2 m
	Tropical Deciduous Forest 3.7 m
	Tropical Evergreen Forest 7.3 m
	Grassland 2.6 m
	Tropical Grassland/Savanna 15 m
	Tundra 0.5 m

g. Evapotranspiration coefficient table (Kc)

Potential Evapotranspiration = ET Coefficient x Reference Evapotranspiration.
ET coefficient values for crops are readily available from irrigation and horticulture handbooks. FAO has an online resource for this. Values for other vegetation can be estimated using Leaf Area Index (LAI) relationships, which is a satellite imagery product derived from NDVI analysis. A typical LAI - ETcoef relationship[8] might look as follows:

$$ETcoef = \begin{cases} \dfrac{LAI}{3} & \text{when } LAI \le 3 \\ 1 & \end{cases}$$

Evapotranspiration coefficients need to be applied to non-vegetated class, such as pavement or water bodies. As a rule of thumb, impermeable surfaces and moving water bodies might be given a low ETcoef value (no zeros should be defined), such as 0.001, to highlight removal of water by drainage. Slow or stagnant water bodies might be given an ETcoef value of 1.
Once evapotranspiration coefficients have been established for all landuse / land classes they must be multiplied by 1000 to obtain the integer value, i.e. Int(ETceof x 1000). No zero values are allowed.

Sample ET coef Table.

ID	Vegetation Type	etk
1	Evergreen Needleleaf Forest	1000
2	Evergreen Broadleaf Forest	1000
3	Deciduous Needleleaf Forest	1000
4	Deciduous Broadleaf Forest	1000
5	Mixed Cover	1000
6	Woodland	1000
7	Wooded Grassland	1000
8	Closed Shrubland	398
9	Open Shrubland	398
10	Grassland	650
11	Cropland (row Crops)	650

8 See SWAT (Soil Water Assessment Tool) manual.

12	Bare Ground	1
13	Urban and Built-Up	1
14	Wetland	1000
15	Mixed evergreen	1000
16	Mixed Forest	1000
17	Orchards/Vineyards	700
18	Pasture	850
19	Sclerophyllous Forests	1000

h. 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-wist.asp, and USGS - http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/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 were filled.

i. Consumptive water use

The consumptive water use for each land use / land class type should be estimated based on agricultural, forestry, and hydrology literature and / or consultation with local professionals in these fields. The value used in the table is an average for each land use type. For crops, water use can be calculated using information on crop water requirements and scaling up based on area covered by crops. In more general agricultural areas, water use by cattle, agricultural processing, etc. must be considered. For forestry, a similar calculation can be made based on estimates of water use by different forest types. 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 must also be considered where applicable. For all of these calculations, it is assumed that the crops, trees, people, etc. are spread evenly across each land use class.

j. Hydropower Watersheds and Sub-watersheds

. See the Working with the DEM section of this manual for information on generating watersheds and sub-watersheds.

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 sub-watersheds, we have provided the global dataset from Hydro1k (http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30/hydro). This dataset can be downloaded from the InVEST distribution server (http://invest.ecoinformatics.org).

k. 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://crunch.tec.army.mil/nidpublic/webpages/nid.cfm). 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/dataweb/). Similar information may be found online at other websites for reservoirs owned or operated by other government agencies or energy companies.

- **g.** *Calibration*: For calibration, data are needed on how much water actually reaches each 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. The model user should consider whether the gage measures natural or managed streamflow and adjust measurements as necessary. The drainage area downstream of the gage and upstream of the hydropower station must also be considered when comparing gaged flow with modeled flow.
- **h.** *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.
- **i.** *Discount_rate*: this rate is defined as how much value the currency loses per year.

l. Seasonality factor (Zhang constant)

The Zhang constant is used to characterize the seasonality of precipitation in the study area, with possible values ranging from 1 to 10. The values are assigned according to the timing of the majority of rainfall in a year. If rainfall primarily occurs during the winter months, Zhang values should be closer to 10; if most rainfall occurs during the summer months or is more evenly spread out during the year, Zhang values should be closer to 1. Our initial testing efforts of this model in different watersheds in different eco-regions worldwide show that this factor is around 4 in tropical watersheds, 9 in temperate watersheds and 1 in monsoon watersheds.

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 sub-basin level. The model then sums and averages water yield at the sub-watershed level. A first run model calibration should be performed using 10 year average input data. For example, if water yield model simulations are being performed under a 1990 land use scenario; climate data (total precipitation and potential evapotranspiration) from 1985 to 1995 should be averaged and used with the 1990 land use map. The other inputs, soil depth and plant available water content are less susceptible to temporal variability and such any available data for these parameters may be used. Observed flow data should be collected from a station furthest downstream in the watershed. As with the climate data, a 10 year average should be used for model calibration. Gauge data is often provided in flow units (i.e m3/s). Since the model calculates water depth, the observed flow data should be divided by the contributing watershed area and converted into units of mm/year. Note, to ensure accuracy, the watershed input being used in the water yield model should have the same approximate area as the contributing watershed area provided with the observed flow data. When assessing the overall accuracy of the model, the mean water yield for the watershed should be compared with the observed depth at the outlet. In nested watersheds or adjacent watershed, calibration could be carried out on one or two stations (watersheds) and validation of these calibrated watersheds could be carried on the other watershed(s).

Before the user starts the calibration process, we highly recommended sensitivity analysis using the observed runoff data. The sensitivity analysis will define the parameters that influence model outputs the most. The calibration can then focus on the highly sensitive parameters followed by the low sensitive ones.

As with all models, model uncertainty is inherent and must be considered when analyzing results for decision making. The model is therefore essentially driven more by parameter values (Z, Kc, root depth) then by the individual physical hydrologic processes taking place in the watershed. Since these parameter values are often obtained from literature or experimental studies under varied conditions, a range of values are usually available (see data sources). InVEST Water Yield model uncertainty is best addressed by performing model simulations under maximum, minimum and mean parameter values. Doing so will provide a range of outputs corresponding to plausible actual conditions.

References

Budyko, M.I. 1974, *Climate and Life,* Academic, San Diego, California.

Donohue, R.J., Roderick, M.L. & McVicar, T.R. 2007, "On the importance of including vegetation dynamics in Budyko's hydrological model.", *Hydrology and Earth System Sciences,* vol. 11, pp. 983-995.

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.

Milly, P.C.D. 1994, "Climate, soil water storage, and the average annual water balance.", *Water Resources Research,* vol. 3, no. 7, pp. 2143-2156.

Potter, N.J., Zhang, L., Milly, P.C.D., McMahon, T.A. & Jakeman, A.J. 2005, "Effects of rainfall seasonality and soil moisture capacity on mean annual water balance for Australian catchments.", *Water Resources Research,* vol. 41.

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.

Zhang, L., Dawes, W.R. & Walker, G.R. 2001, "Response of mean annual evapotranspiration to vegetation changes at catchment scale. ", *Water Resources Research,* vol. 37, pp. 701-708.

WATER PURIFICATION: NUTRIENT RETENTION

Summary

Water purification is an essential service provided by ecosystems. InVEST estimates the contribution of vegetation and soil to purifying water through the removal of nutrient pollutants from runoff. The biophysical model uses data on water yield, land use and land cover, nutrient loading and filtration rates and water quality standards (if they exist) to determine nutrient retention capacity for current and future land use scenarios. The valuation model uses data on water treatment costs and a discount rate to determine the value contributed by the natural system to water purification. It does not address chemical or biological interactions besides filtration by terrestrial vegetation (such as in-stream processes) and is less relevant to locations with extensive tile drainage or ditching, strong surface water-ground water interactions, or hydrology dominated by infiltration excess (dry regions with flashy rains)..

Introduction

Clean water is a vital service provided by healthy streams, watersheds and river basins. Polluted water is especially harmful to human health. In fact, waterborne illnesses are the leading cause of human disease and death around the world killing more than 3.4 million people annually (World Health Organization). Clean water also provides habitat for aquatic life in streams, rivers and lakes but these habitats require a proper nutrient balance. If nutrients and toxins accumulate in water, fish and other aquatic creatures may be poisoned, along with the people consuming them.

Many of these harmful conditions are caused by non-point source pollution, which occurs when a pollution source is distributed over an area or discharged into the atmosphere and incorporated into hydrological flows through rainfall and runoff. There are numerous sources of non-point source pollution, including fertilizer used in agriculture and residential landscaping, and oil that leaks from cars onto roads. When it rains or snows, water flows over the landscape carrying pollutants from these surfaces into streams, rivers, lakes, and the ocean.

One way to reduce non-point source pollution is to reduce the amount of pollutants that enter the water body. If this is not possible, ecosystems can provide this service by retaining some non-point source pollutants. 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 defense against pollutants entering a stream.

Land use planners from government agencies to environment groups need information regarding the contribution of ecosystems to mitigating water pollution. Specifically, they require information pertaining to the value of every part of a watershed for maintaining water quality so that conservation may be targeted to the areas most important for protecting a safe water supply for people and aquatic life. They can also use this information to avoid impacts in areas that currently contribute the most to filtering out pollutants. The InVEST Tier 1 model provides this information for non-point source pollutants. We have designed the model to deal with nutrient pollutants (nitrogen and phosphorous), but the model can be used for other kinds of contaminants (persistent organics, pathogens etc.) if data are available on the loading rates and filtration rates of the pollutant of interest.

The Model

The InVEST Water Purification Nutrient Retention model calculates the amount of nutrients retained on every pixel then sums and averages nutrients export and retention per sub-watershed. 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 sub-watershed to watershed scale. We are only confident in the interpretation of these models at the sub-watershed scale, so all outputs are summed and/or averaged to the sub-basin 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.**

InVEST also calculates the economic value that nutrient retention provides through avoided treatment costs. It integrates data on the magnitude of overland flow, pollutant loading, the capacity of different vegetation types to filter pollutants, the cost of water treatment (for pollutants of interest), and feasibility to meet water quality standards.

The model's limitations are that it runs on an annual average basis, can only assess one pollutant per run, does not address chemical or biological interactions besides filtration by terrestrial vegetation, and in some cases it may provide an in-accurate marginal cost for pollutant removal (when pollutant costs relative to pollutant concentration are non-linear). The model assumes that non-point sources of water pollution result from export that can be mitigated by vegetation serving as intercepting filters. It also assumes that water flows downslope along natural flowpaths, so it may be less relevant in areas with tile drainage and extensive ditching practices. It does not consider the role of ecosystems in affecting point-source pollutants. It also may be less relevant where there is significant groundwater surface water interaction and in dry eco-regions

How it Works

The model runs on a gridded map. It estimates the quantity and value of pollutants retained for water purification from a landscape in three components.

The first step calculates annual average runoff from each parcel (see Hydropower chapter).
In the second step, we determine the quantity of pollutant retained by each parcel on the landscape. First, we estimate how much pollutant is exported from each parcel, based on export coefficients the user inputs. Export coefficients, developed by Reckhow et al. 1980, are annual averages of pollutant fluxes derived from various field studies that measure export from parcels within the United States. Since these coefficients are averages fluxes , we include a hydrological sensitivity score that accounts for differences in condition between the fields where the measures were developed and the conditions where the user is applying the model. We do this with the following equation:

$$ALV_x = HSS_x \cdot pol_x$$

where ALV_x is the Adjusted Loading Value at pixel x, pol_x is the export coefficient at pixel x, and HSS_x is the Hydrologic Sensitivity Score at pixel x which is calculated as:

$$HSS_x = \frac{\lambda_x}{\overline{\lambda}_W}$$

where λ_x is the runoff index at pixel x, calculated using the following equation, and $\overline{\lambda}_W$ is the mean runoff index in the watershed of interest.

$$\lambda_x = Log\left(\sum_U Y_u\right)$$

where $\sum_U Y_u$ is the sum of the water yield of pixels along the flow path above pixel x (it also includes the water yield of pixel x).

Once we know how much pollutant leaves each parcel, we can determine how much of that load is retained by each downstream parcel, as surface runoff moves the pollutant towards the stream. The model routes water down flow paths determined by slope, and allows each parcel downstream from a polluting parcel to retain pollutant based on its land cover type and that land cover type's ability to retain the modeled pollutant. We do not account for saturation of uptake. By following the pollutant load of each parcel all the way downstream to a water body, the model also tracks how much pollutant reaches the stream. The table below describes how this removal from routing and hydraulic connectivity is done:

Cell	Vegetation retention	ALV	Retained by Cell (retained)	Outflow quantity (OQ) from Cell (Gi=1-Ei)

1	E1	ALV1	0	ALV1
2	E2	ALV2	ALV1.E2	ALV1.G2+ALV2
3	E3	ALV3	((ALV1.G2+ALV2).E3	(ALV1.G2+ALV2).G3+ALV3
4	E4	ALV4	ALV1.G2.G3.E4+ ALV2.G3.E4+ ALV3.E4	ALV1.G2.G3.G4+ ALV2.G3.G4+ ALV3.G4+ ALV4

The model then aggregates the loading that reaches the stream from each pixel to the sbu-watershed then to the watershed level. The user can then compare this load (adding the point sources loadings if any) to a known (observed or simulated using another water quality model) measurement and adjust export coefficients and removal efficiencies (vegetation retention) as needed until the modeled load matches the measured load for each point of interest. The user should consider the likely impact of in-stream processes in any calibration work as this model does not include in-stream processes.

Once the total load is determined, we can (optionally) calculate the value of this service provided by each pixel based on the avoided treatment costs the retention by natural vegetation and soil provides. We make this calculation as follows:

$$wp\_Value_x = Cost(p) * retained_x * (1 - \frac{Ann\_Load}{CNL})$$

Where:
Wp_value$_x$ is the value of retention for pixel x.
Cost(p) is the annual treatment cost in $(currency)/kg for the pollutant of interest (p).
retained$_x$ is the total pollutant retained by pixel x of the total amount of upslope pollutant arriving at this pixel.
Ann_Load is the Annual Loading for pollutant of interest allowed before any damage to human health or natural habitat health occur. It is the threshold below which service is invaluable.
CNL is the cumulative pollutant loading estimated by the model at the outlet of your watershed (outlet of your watershed = point of interest)

This pixel value is then summed to the sub-watershed to determine the water purification value per sub-watershed.

Limitations and Simplifications

The model has a number of assumptions. First, since the model was developed for watershed and landscapes dominated by saturation excess runoff hydrology, it may be less applicable to locations where the hydrology is determined by rainfall intensity; in areas where flashy rains are predominant and where infiltration excess runoff occurs. This kind of runoff is the result of intense rains that saturate only the top soil layer, not the entire profile. However, the model's use of a runoff index and hydraulic routing should sufficiently adjust for this.

Second, the model can only assess one pollutant per run. If the user wishes to model several pollutants, but does not have data on loadings and filtration rates for each pollutant, choose a pollutant that acts as a surrogate in predicting loadings for other pollutants. The most common surrogate is phosphorus because heavy phosphorus loadings are often associated with other pollutants such as nitrogen, bacteria and suspended solids, however using a pollutant surrogate should be approached with caution. Alternatively, the user can run the model multiple times using export values and retention coefficients for each pollutant. In general, the model can only assess pollutants that are susceptible to export via surface and subsurface flows.

Third, the model does not address any chemical or biological interactions that may occur from the point of loading to the point of interest besides filtration by terrestrial vegetation. In reality, pollutants may degrade over time and distance through interactions with the air, water, other pollutants, bacteria or other actors.

Fourth, the model assumes that there is continuity in the hydraulic flow path. The user should be aware of any discontinuity in the flow path. Tile drainage and ditches could create short cuts for pollutant movement and run pollutant directly to streams

Finally, in some cases the model may provide an inaccurate marginal cost for pollutant removal. The full marginal cost of removing a unit volume of pollutants is difficult to estimate due to the complexity of the treatment process. The marginal cost may not be a constant value but instead a function of decreasing cost per additional unit volume of pollutant as the total volume increases. Also, the cost of treatment may change over time as technology improves or water quality standards evolve.

Data Needs

The model uses eleven types of input data. See the appendix for information on finding or creating these inputs.

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 necessary 'burning' hydrographic features into the elevation model (recommended when you see unusual streams). See the Working with the DEM section of this manual for more information.

Name: File can be named anything, but avoid spaces ("DEM" is simplest).
Format: Standard GIS raster file (e.g., ESRI GRID or IMG), with elevation value for each cell given in meters above sea level.
Sample data set: \InVEST\Base_Data**dem**

2. Soil depth (required). A GIS raster dataset with an average soil depth value for each cell. The soil depth values should be in millimeters .

Name: File name can be anything, but avoid spaces.
Format: Standard GIS raster file, with an average soil depth in millimeters for each cell.
Sample data set: \InVEST\ Base_Data **soil_depth**

3. 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 avoid spaces.
Format: Standard GIS raster file (e.g., ESRI GRID or IMG), with precipitation values for each cell.
Sample data set: \InVEST\ Base_Data **precip**

4. 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.

Name: File can be named anything, but avoid spaces.
Format: Standard GIS raster file (e.g., ESRI GRID or IMG), with available water content values for each cell.
Sample data set: \InVEST\Base_Data**pawc**

5. Average Annual Potential Evapotranspiration (required). A GIS raster dataset, with an annual average evapotranspiration value for each cell. Potential 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 evapotranspiration values should be in millimeters.

Name: File can be named anything, but avoid spaces.
Format: Standard GIS raster file (e.g., ESRI GRID or IMG), with potential evapotranspiration values for each cell.
Sample data set: \InVEST\Base_Data**eto**

6. 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 avoid spaces.
Format: Standard GIS raster file (e.g., ESRI GRID 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 Model Coefficients table.
Sample data set: \InVEST\Base_Data**landuse_90**

7. Sub-watersheds (required). A shapefile of polygons. This is a layer of sub-watersheds, which contribute to the points of interest where water quality will be analyzed. See the Working with the DEM section for information on creating sub-watersheds.

Due to limitations in ArcMap geoprocessing, the maximum size of a sub-watershed is approximately the equivalent of 4000x4000 cells, with cell size equal to the smallest cell size of your input layers. If the whole watershed contributing to a point of interest is larger than this size, it will need to be divided into sub-watersheds that are each smaller. Then, the resulting sub-watershed layer should be entered here and the whole watershed layer should be used in the Watersheds input. If the whole watershed is smaller, then it does not need to be divided, and the same watershed layer should be entered for both Sub-watersheds and Watersheds inputs. Sub-watersheds will be mosaicked back together into whole watersheds for the final output.

Name: File can be named anything, but avoid spaces.
Format: A shapefile of polygons with unique integers for each sub-watershed in the subws_id field.
Sample data set: \InVEST\ Base_Data**subwatersheds.shp**

8. 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. See **Sub-watersheds** above for more information on watershed size requirements.

Name: File can be named anything, but avoid spaces.
Format: Standard GISshapefile , with unique integers values for each watershed in the ws_id field
Sample data set: \InVEST\ Base_Data **watersheds.shp**

9. Model Coefficients Table (required). A table of land use/land cover (LULC) classes, containing data on water quality coefficients used in this tool. NOTE: these data are attributes of each LULC class rather than attributes of individual cells in the raster map.

Name: File can be named anything.
File type: *.dbf
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:

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 above.
b. *LULC_desc*: Descriptive name of land use/land cover class (optional)
c. *root_depth:* The maximum root depth for vegetated land use classes, given in integer millimeters. Non-vegetated LULCs should be given a value of 1.
d. *etk:* The evapotranspiration coefficient for each LULC class, used to obtain actual evapotranspiration by using plant energy/transpiration characteristics to modify the reference evapotranspiration, which is based on alfalfa (or grass). Coefficients should be multiplied by 1000, so that the final *etk* values given in the table are integers ranging between 1 and 1500 (some crops evapotranspire more than alfalfa in some very wet tropical regions and where water is always available).
c. *load_n / load_p*: The nutrient loading for each land use. If nitrogen is being evaluated, supply values in *load_n*, for phosphorus, supply values in *load_p*. The potential for terrestrial loading of water quality impairing constituents is based on nutrient export coefficients. The nutrient loading values are given as integer values and have units of

g. Ha^{-1} yr^{-1}.

d. *eff_n / eff_p*. The vegetation filtering value per pixel size for each LULC class, as an integer percent between zero and 100. If nitrogen is being evaluated, supply values in *eff_n*, for phosphorus, supply values in *eff_p*. This field identifies the capacity of vegetation to retain nutrient, as a percentage of the amount of nutrient flowing into a cell from upslope. For example if the user has data describing that wetland of 5000 m^2 retains 82% of nitrogen, then the retention efficiency that the he should input into this filed for eef_n is equal to (82/5000 * (cell size)^2). In the simplest case, when data for each LULC type are not available, high values (60 to 80) may be assigned to all natural vegetation types (such as forests, natural pastures, wetlands, or prairie), indicating that 60-80% of nutrient is retained. An intermediary value also may be assigned to features such as contour buffers. All LULC classes that have no filtering capacity, such as pavement, can be assigned a value of zero.

Sample data set: \InVEST\Base_Data\Water_Tables.mdb**Biophysical_Models**

Example: Case with 6 LULC categories, where potential evapotranspiration, root depth and nutrient (both N and P) filtration efficiencies do not vary among LULC categories, while nutrient loadings do.

LULC_desc	lucode	etk	root_depth	load_n	eff_n	load_p	eff_p
Low Density Residential	1	1	1	7000	0	1000	0
Mid Density Residential	2	1	1	7250	0	1100	0
High Density Residential	3	1	1	7500	0	1200	0
Very High Density Residential	4	1	1	7750	0	1300	0
Vacant	5	1	1	4000	0	100	0
Commercial	6	1	1	13800	0	3000	0

10. Threshold flow accumulation value (required). Integer value defining the number of upstream cells that must flow into a cell before it's considered part of a stream. This is used to generate a stream layer from the DEM. The default is 1000. This value is to define and create the stream lines. If the user has map of stream lines in the watershed of interest, he/she should compare it the V_stream map (output of the model). This value also needs to be well estimated in watersheds where tile drainage and ditches are present. This threshold expresses where hydraulic routing is discontinued and where retention stops and the remaining of the pollutant will be exported to the stream.

11. Water Purification Valuation table. This is a table containing calibration and valuation information for each of the points of interest. There must be one row for each watershed in the Watersheds layer.

Name: File can be named anything.
File type: *.dbf
Rows: Each row corresponds to a watershed.
Columns: Each column contains a different attribute of each watershed and must be named as follows:

a. *ws_id* (watershed ID): Unique integer value for each watershed, which must correspond to values in the Watersheds layer.
b. *calib*: Annual watershed loading calibration constant. Multiplying this value by the total watershed load output (Output/ws_nutexp) should give the actual total annual load observed/measured at the point of interest. Floating point value.
c. *ann_load*: The total critical annual nutrient loading allowed for the nutrient of interest at the point of interest. Floating point value. It has unit of $Kg.yr^{-1}$.
d. *cost*: Annual Cost of nutrient removal treatment in $ / kg removed. Floating point value.
e. *time_span:* Number of years for which net present value will be calculated. Integer value. This could be the time span (number of years) of either the same LULC scenario or the water treatment plant life span.
f. *discount:* The rate of discount over the time span, used in net present value calculations. Floating point value.

Sample data set: \InVEST\Base_Data\Water_Tables.mdb**Water_Purification_Valuation**

ws_id	calib	ann_load	cost	time_span	discount
0	1	50	24	15	5
1	1	77	24	25	5
2	1	31	24	15	5

12. Water Purification threshold table. A table containing annual nutrient load threshold information for each of the points of interest. There must be one row for each watershed in the Watersheds layer.

Name: File can be named anything.
File type: .dbf or .mdb
Rows: Each row corresponds to a watershed.
Columns: Each column contains a different attribute of each watershed and must be named as follows:

a. *ws_id* (watershed ID): Unique integer value for each watershed, which must correspond to values in the Watersheds layer.
b. *ann_load*: The total critical annual nutrient loading allowed for the nutrient of interest at the point of interest. Floating point value. It has unit of $Kg.yr^{-1}$.

Sample data set: C: \Invest\Base_Data\Water_Tables.mdb**Water_Purification_Threshold**

Running the Model

Before running the Water Purification Nutrient Retention model, make sure that the INVEST toolbox has been added to your ARCMAP document, as described in the **Getting Started** chapter of this guide. Second, make sure that you have prepared the required input data files according to the specifications in *Data Needs*.

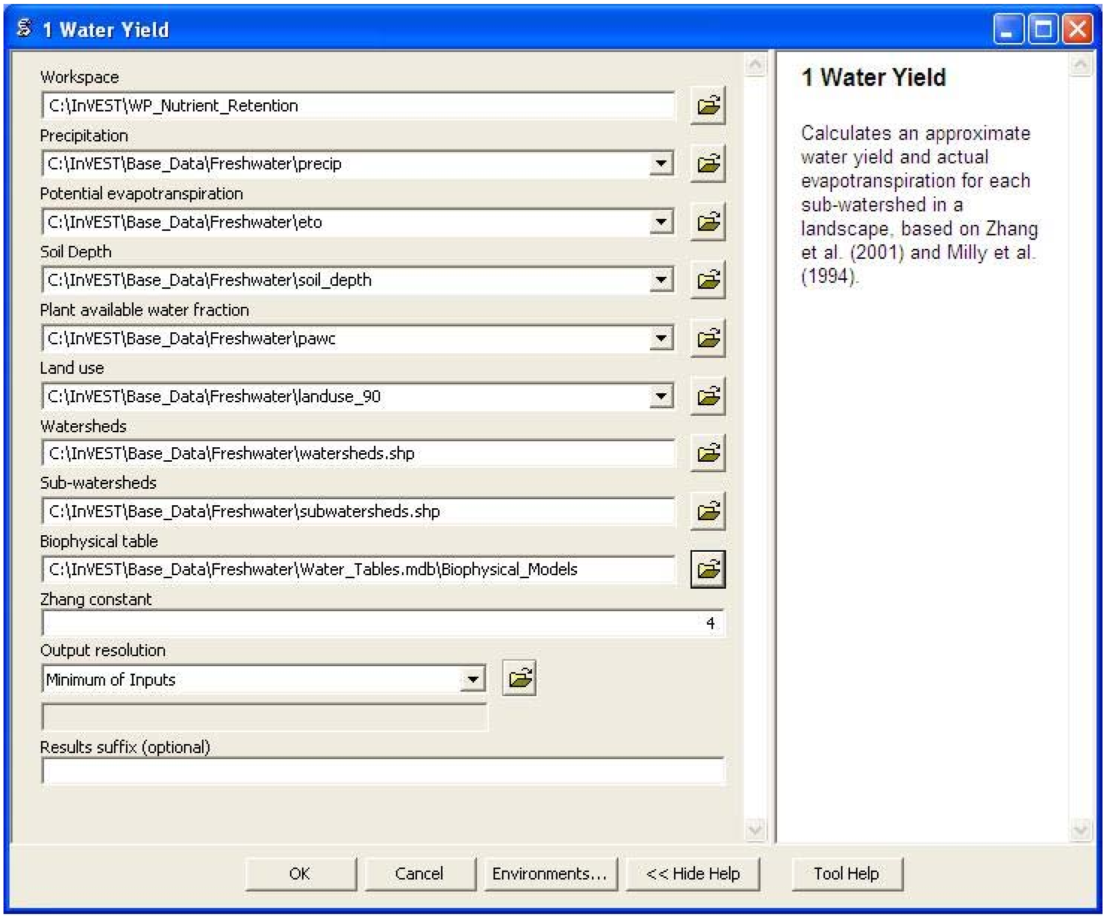
- Create a workspace on your computer hard-drive if you are using your data. The pathname to the workspace should not have spaces. All your output files will be saved here. For simplicity, you could create a folder in your workspace called "Input" and place all your input files here. It is not necessary to place input files in the workspace, but this will make it easier to view the data you use to run your model. If this is your first time using InVEST and you wish to use sample data, you can use the data provided in InVEST-Setup.exe. If you unzipped the InVEST files to your C-drive (as described in the **Getting Started** chapter), you should see a folder named /InVEST/ **WP_Nutrient_Retention**. This folder will be your workspace. The input files are in /InVEST/Base_Data**/Freshwater/**.

- Open an ARCMAP document to run the model.

- Locate the INVEST toolbox in ARCTOOLBOX. ARCTOOLBOX should be open in ARCMAP, but if it is not, click on the ARCTOOLBOX symbol. See the **Getting Started** chapter if you do not see the InVEST toolbox.

- Click the plus sign on the left side of the INVEST toolbox to expand the list of tools. Double-click on Nutrient_Retention. Three options will appear: Water Yield, Nutrient Removal, and Valuation. Water Yield must be run first, Nutrient Removal second, and Valuation last. The scripts MUST be run in this order because the output from a previous script is required for the next script.

- Click on Water Yield.



- An interface will appear like the one above that indicates default (sample data) file names, but you can use the file buttons or drop-down arrows to browse to your data. When you place your cursor in each space, you can read a brief description of the data requirements in the right side of the interface. Refer to the ***Data Needs*** section for information on data formats.

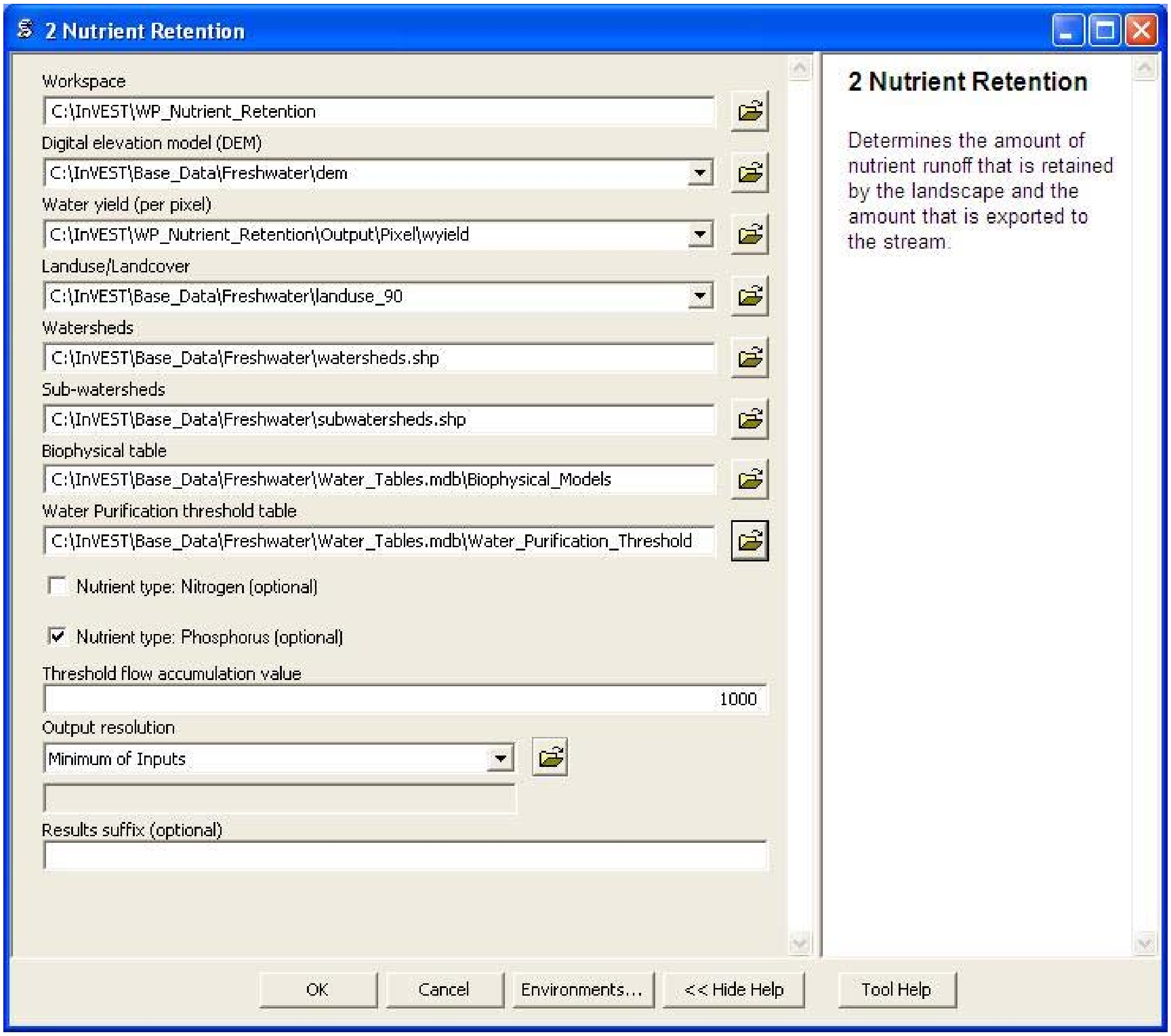
- Fill in data file names and values for all required prompts. Unless the space is indicated as optional, it requires data.

- After entering all required data, click OK. The script will run, and its progress will be indicated by a "Progress dialogue".

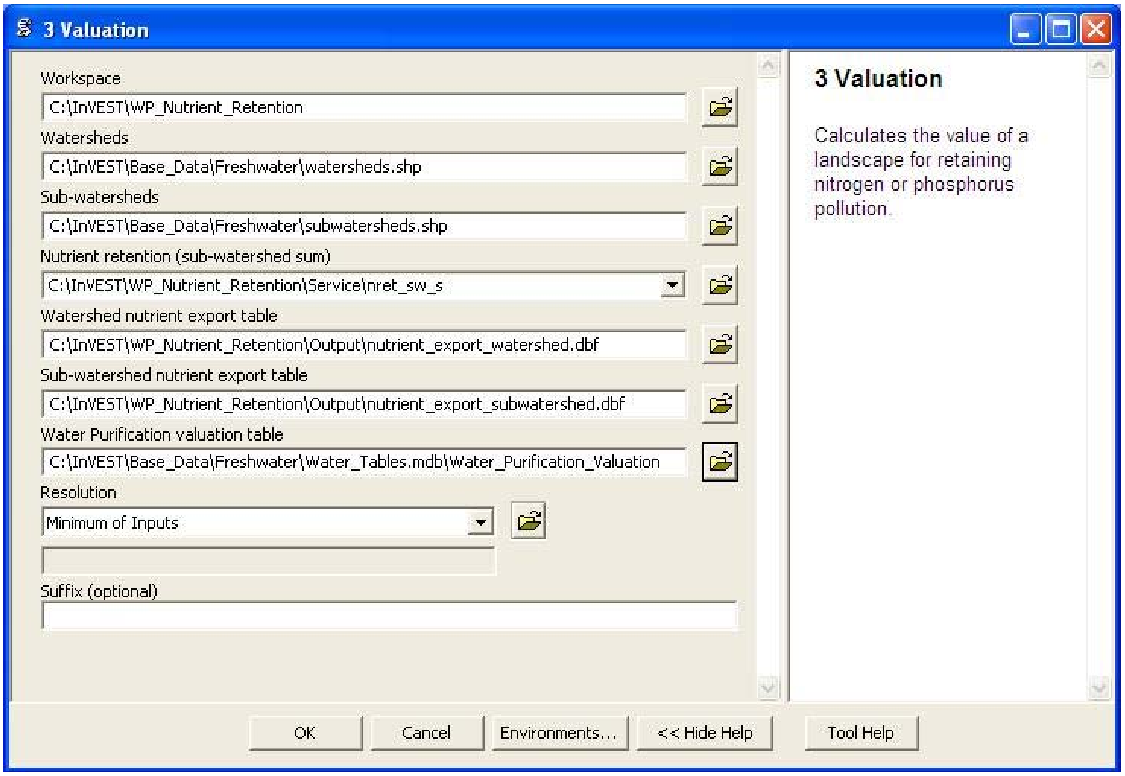
- To view the attribute data of output files, right click a layer and select OPEN ATTRIBUTE TABLE.

- Now you are ready to run Nutrient Removal. Follow the same steps as for Water Yield. Note that an output from Water Yield, **Output\Pixel\wyield**, is a required input to Nutrient Retention. **Make sure to select one of the Nutrient Type boxes, the model needs one of the two to be checked to run** (sometimes, you will see (optional) after Nitrogen or

Phosphorus, but you still need to check the box of the nutrient you are interested in). The interface is below:



- When the script completes running, its results will be saved in the Output and Service folders.
- Load the output grids into ARCMAP using the ADD DATA button.
- Finally, you have the option to run Valuation. Three outputs from Nutrient Removal are required, Service\nret_sw_s, Output\nutrient_export_watershed.dbf, and Output\nutrient_export_subwateshed.dbf.. The interface is below:



- When the script completes running, its results will be saved in the Service folder.
- Load the output grids into ARCMAP using the ADD DATA button.
- To view the attribute data of output files, right click a layer and select OPEN ATTRIBUTE TABLE.

Interpreting Results

Parameter Log

Each time the model is run, a text 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.

Final Results

Final results are found in the Output and Service folders within the working directory set up for this model.

Output**adjl_sw_m** (kg/ha): Mean adjusted load per sub-watershed.

- Output**adjl_sw_s** (kg): Total adjusted load per sub-watershed.

- Service**nret_sw_s** (kg): Total amount of nutrient retained by each sub-watershed.

- Service**nret_sw_mn** (kg/ha): Mean amount of nutrient retained by each sub-watershed.

- Output**nexp_sw_m** (kg/ha): Raster showing the mean amount of nutrient per sub-watershed that is exported to the stream.

- Output**nexp_sw_s** (kg): Raster showing the total amount of nutrient per sub-watershed that is exported to the stream.

- *Output\nutrient_export_watershed.dbf:* Table containing values for the total nutrient loading for each watershed
- Output**nutrient_export_subwatershed.dbf**: Table containing values for the total nutrient loading for each sub-watershed.

- Output**nutrient_export_watershed.dbf**: Table containing values for the total nutrient loading for each watershed.

- Service**nut_val** ($, currency): This grid shows the economic **b**enefit per sub-watershed of filtration by vegetation delivered at the downstream point(s) of interest. **THIS OUTPUT REPRESENTS THE ECOSYSTEM SERVICE OF WATER PURIFICATION IN ECONOMIC TERMS.** It may be useful for identifying areas where investments in protecting this ecosystem service will provide the greatest returns. Variation in this output with scenario analyses (by running and comparing different LULC scenarios) will indicate where land use changes may have the greatest impacts on service provision.

- Service**nutrient_value_watershed.dbf:** Table containing values for the total nutrient retention per watershed, and the value of that retention for keeping nutrient from arriving at the watershed outlet point of interest.

- Service**nutrient_value_subwatershed.dbf:** Table containing values for the total nutrient retention per sub-watershed, and the value of that retention for keeping nutrient from arriving at the watershed outlet point of interest.

These outputs provide an interim insight into the dynamics of pollutant loading, transport and filtration in a watershed. The model will be most informative if it is used in collaboration with experts in hydrology familiar with the watershed. In case model coefficients require adjustment and to guard against erroneous data input, it is recommended that model outputs are verified with field data mimicking pollutant loading and watershed transport processes.

Appendix: 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. It will be updated as new data sources and methods become available.

In general, the FAO Geonetwork could 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. Free raw global DEM data is available on the internet from the World Wildlife Fund - http://www.worldwildlife.org/freshwater/hydrosheds.cfm. NASA provides free global 30m DEM data at http://asterweb.jpl.nasa.gov/gdem-wist.asp. Or, it may be purchased relatively inexpensively at sites such as MapMart (www.mapmart.com). The hydrological aspects of the DEM used in the model must be correct. Please see the introduction section to the complete model manual for instructions on how to check for this.

2. Soil depth.
Soil depth may be obtained from standard soil maps. Coarse, yet free global soil characteristic data are available at http://www.ngdc.noaa.gov/seg/cdroms/reynolds/reynolds/reynolds.htm. The FAO also provides global soil data in their Harmonized World Soil Database: http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/.
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://soils.usda.gov/survey/geography/ssurgo/ and STATSGO http://soils.usda.gov/survey/geography/statsgo/ . 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 depth should be calculated as the max depth of all horizons within a soil class component, and then a weighted average of the components should be estimated. This can be a tricky GIS analysis: In the US soil categories, each soil property polygon can contain a number of soil type components with unique properties, and each component may have different soil horizon layers, also with unique properties. Processing requires careful weighting across components and horizons. The Soil Data Viewer (http://soildataviewer.nrcs.usda.gov/), a free ArcMap extension from the NRCS, does this soil data processing for the user and should be used whenever possible.

Ultimately, a grid layer must be produced. Data gaps, such as urban areas or water bodies need to be given appropriate values. Urban areas and water bodies can be thought of having zero soil depth.

4. Land use and land cover
A key component for all water models is a spatially continuous land use and land cover raster grid.

That is, within a watershed, all land use and land cover 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. The more detailed and descriptive these files are the better accuracy and modeling results. Global land use data is available from the University of Maryland's Global Land Cover Facility: http://glcf.umiacs.umd.edu/data/landcover/. This data is available in 1 degree, 8km and 1km resolutions. 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 could involve 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. The user should only break up a land use type if it will provide more accuracy in modeling. For instance, for the Water Purification: Nutrient 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.

5. Nutrient Loading Coefficients

Examples of export and loading coefficients can be found in the EPA PLOAD User's Manual http://www.epa.gov/waterscience/basins/b3docs/PLOAD_v3.pdf and in the Wetlands Regulatory Assistance Program publication http://el.erdc.usace.army.mil/elpubs/pdf/tnwrap043.pdf. Note that the examples in the EPA guide are in lbs/ac/yr and would need to be converted to kg/ha/yr.

Phosphorus is a common water quality proxy because it incorporates both dissolved and particulate nutrient loadings, are well associated with surface runoff, and is usually the limiting nutrient for fresh water systems. The table below shows default phosphorus export coefficients largely based on values from USEPA manuals, and research studies in the US. The bottom three rows are used solely for direct untreated waste water discharge (i.e. untreated sewage piped into water systems) from urban areas commonly found in developing countries.

If local data / approximations for Phosphorus export coefficients exist they can be used or replace default values in table.

Table : Example Phosphorus and Nitrogen export coefficients (Reckhow et al. 1980)

Landuse	Nitrogen Export Coefficient (kg/ha/yr)	Phosphorus Export Coefficient (kg/ha/yr)
Forest	1.8	0.011
Corn	11.1	2

Cotton	10	4.3
Soybeans	12.5	4.6
Small Grain	5.3	1.5
Pasture	3.1	0.1
Feedlot or Dairy	2900	220
Idle	3.4	0.1
Residential	7.5	1.2
Business	13.8	3
Industrial	4.4	3.8

The loading proxy may also aggregate several indicators, agreed upon between managers, such as an algorithm that aggregates phosphorus, nitrates, and other constituents. Alternatively, a manager may begin using values from EPA table as a starting point to generate discussion, and then alter values based on local expert opinion and stakeholder feedback.

6. Removal Efficiencies (*eff_n, eff_p*)

These values are used to incorporate the effects of natural vegetation that buffer potential water quality impairment downhill from sources. To develop these values, all land class pixels that contain natural vegetation (such as forests, natural pastures, wetlands, or prairie) are assigned high values and vegetation that has no or little filtering value receives a value of zero. All values should fall between 0 and 100. Consult with a hydrologist if not certain about assignment of specific values.

7. Calibration Data (*Calib*).

Calibration data is needed for ensuring that the Tier 1 Water Purification: Nutrient Retention model results match well with reality. Most often calibration data may be obtained from water quality monitoring that is already in place. If the point of interest is a water supply intake, the drinking water entity will most likely collect water quality at the point of intake. If the point of interest is in a stream or lake, the water quality may have been tested by a public agency. Most likely if the location is of interest in terms of meeting a water quality standard, data should be available. In the U.S. the user may contact or look up online their state environmental agency, EPA, fish and wildlife service, or any local universities conducting research on the water body.

Once data is collected, the user may have to convert the values into actual pollutant loads and / or correlate a measured pollutant with a proxy modeled pollutant. In addition to correlation analysis, other calibration methods such as Nash Coefficient, ranking analysis, and graphical comparison could be used.

8. Critical Annual Load (*Ann_Load*).

Gathering information on water quality standards or targets should be part of the formulation of modeling objectives. If the target to be met is a drinking water target, standards may be set by the federal, state or local level (whichever standard is the most stringent). The table below provides some general drinking water standards set by global and national agencies.

Selected Drinking Water Standards by World Health Organization, European Union, and US

EPA. (Ashbolt et al. 2001)

Parameter	Type	WHO	EU	USEPA
Ammonia	Social	1.5 mg L -1	0.50 mg L -1	No GL
pH	Social	6.5-8	No guidelines	6.5-8.5
Chloride	Social	250 mg L -1	250 mg L -1	250 mg L -1
Iron	Social	0.3 mg L -1	0.2 mg L -1	0.3 mg L -1
Lead	Health	0.01 mg L -1	0.01 mg L -1	0.015 mg L -1
Arsenic	Health	0.01 mg L -1	0.01 mg L -1	0.01 mg L -1
Copper	Health	2.0 mg L -1	2.0 mg L -1	1.3 mg L -1
Fecal Coliform bacteria	Health	0 counts/100 mL	0 counts/100 mL	0 counts/100 mL

These standards are set for point of use, meaning that the standard at the point of interest, where water supply will be drawn, may be more relaxed than these standards if water treatment is in place. In-situ water quality standards (for rivers, lakes and streams) may also be set at the national, state and local level. They may be the same across all water bodies of the same type (in rivers, for example) or they may vary depending on the established use of the water body or the presence of endangered species. In the U.S. Total Maximum Daily Loads of various pollutants are typically established by state regulatory agencies in compliance with the Clean Water Act. States report information on TMDL's to the U.S.EPA on specific waterways http://www2.ctic.purdue.edu/kyw/tmdl/statetmdllists.html .

9. Marginal pollutant removal costs (*Cost*)

The cost to remove pollutants may vary greatly for each point of interest. If the point of interest is a water supply outtake, this value should be obtained from the water treatment entity who uses and treats the water. Calculations may need to be performed to transform actual costs to cost per unit volume of pollutant, and correlations may need to be run between a proxy pollutant and other pollutants that the treatment process removes. If a more general cost of treatment is sought, the user may consult engineering texts or literature to obtain average costs. The user must be sure to bring these costs into present value and make adjustments as necessary depending on the location and type of treatment.

If the point of interest is an in situ water quality target, the marginal pollutant removal cost is much more difficult to obtain. The user may be able to estimate the cost of an additional unit volume of pollutant in terms of fish populations, lost revenue for recreation, or a fine, but this may be a complicated calculation not worth the effort at this level of modeling. The user may choose to assign a cost of one to save time while still obtaining relative results useful in comparing scenarios.

References

Anderson, J. R., et al. 1976. A Land Use And Land Cover Classification System For Use with Remote Sensor Data: Geological Survey Professional Paper 964. Edited by NJDEP, OIRM, BGIA, 1998, 2000, 2001, 2002, 2005.

Ashbolt, N.J., Grabow, W.O.K. and Snozzi, M. 2001. "Indicators of microbial water quality," in Water Quality: Guidelines, Standards and Health, L. Fretwell and J. Bartram, Editors. 2001, World Health Organization (WHO), IWA Publishing: London, U.K.

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.*, U.S. Environmental Protection Agency, Washington, D.C.

Uusi Kamppa, J., E. Turtola, H. Hartikainen, T. Ylaranta. 1997. The interactions of buffer zones and phosphorous runoff. In Buffer zones: Their processes and potential in water protection, eds.

N. Haycock, T. Burt, K. Goulding, and G. Pinay, 43–53. Hertfordshire, UK: Quest Environmental.

SEDIMENT RETENTION MODEL

AVOIDED DREDGING AND WATER QUALITY REGULATION

Summary

Reservoirs are linked to a number of ecosystem services, including the generation of energy through reservoir hydropower production, irrigation of crops and recreational activities. Erosion and sedimentation of watersheds can lead to decreased hydropower output, structural damage to reservoirs and other water infrastructure, and flooding. InVEST estimates the capacity of a land parcel to retain sediment using data on geomorphology, climate, vegetation and management practices. These estimates are combined with data on sediment removal costs, reservoir design, and a discount rate to calculate the avoided cost of sediment removal. Limitations of the model include negligence of mass erosion, inadequate information about sediment removal costs, and simplified LULC classifications.

Introduction

Erosion and sedimentation are natural processes that contribute to healthy ecosystems, but too much may have severe consequences. Excessive erosion can reduce agricultural productivity, increase flooding and pollutant transport, and threaten bridges, railroads and power infrastructures. Erosion can lead to sediment build-up, which strains water infrastructures, such as reservoirs and flood control systems, and increases water treatment costs. Sedimentation is particularly problematic for reservoirs, which are designed to retain sediment as water is released. Regular sediment removal can avoid some of these issues but this involves expensive maintenance costs.

The magnitude of sediment transport in a watershed is determined by several factors. Natural variation in soil properties, precipitation patterns, and slope create patterns of erosion and sediment runoff. Vegetation holds soil in place and captures sediment moving overland. However, changes in land management practices can alter the sediment retention capacity of land by removing important vegetation.

There are many clear examples of the effects of LULC change on erosion and sedimentation. Forest fires that clear significant areas of vegetation are often followed by mudslides when heavy rains occur (Meyer *et al*. 2001). After the fire the vegetation that once held sediment in place no longer exists and the top layers of soil can be carried downstream by overland runoff. Deforestation results in a similar process, although in some cases it may occur on longer time scales. Even in areas where land cover remains the same, a change in land use practice can alter the sediment retention capacity of the landscape. For example, moving from no-till to till agriculture has been shown to increase the rate of soil erosion. The continuous accumulation of increased sediment loads as a result of changes in LULC can cause serious problems such as increasing siltation rate, and increasing dredging costs that were not anticipated during the original design of reservoir infrastructure, maintenance and operation plans.

To reduce the damages and costs associated with sedimentation, land, water and reservoir managers require information regarding the extent to which different parts of a landscape contribute to sediment retention, and how land use changes may affect this retention. Such information can support decisions by government agencies, businesses, and NGOs. For example, a power company operating a hydropower reservoir may elect to conserve upstream forests that maintain a sediment retention service if the cost of conserving the forests is less than the costs of reduced hydropower potential, sediment removal, and dam replacement. Maps showing which forest parcels offer the greatest sediment retention benefits would help the power company maximize returns on their investment. InVEST aims to provide these kinds of information.The outputs from these models will allow planners and managers to consider how LULC change in one area in the watershed can cause sedimentation problems at other locations.

The Model

The Sediment Retention model provides the user with a tool for calculating the average annual soil loss from each sub-watershed of land, determining how much of that soil may arrive at a particular point of interest, estimating the ability of each parcel to retain sediment, and assessing the cost of removing the accumulated sediment on an annual basis. An important determinant of soil retention capacity is land use and land cover. To identify a land parcel's potential soil loss and sediment transport, the InVEST Avoided Reservoir Sedimentation model uses the Universal Soil Loss Equation (USLE) (Wischmeier & Smith 1978) at the pixel scale, which integrates information on LULC patterns and soil properties, as well as a digital elevation model, rainfall and climate data. 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 sub-watershed to watershed scale. We are only confident in the interpretation of these models at the sub-watershed scale, so all outputs are summed and/or averaged to the sub-basin 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.**

This model can also be used to value the landscape vis-à-vis maintaining water quality or avoiding reservoir sedimentation. In the water quality maintenance case, the model uses additional information on water quality standards and treatment costs to value the ability of each sub-watershed to reduce treatment costs. In the reservoir maintenance case, the model uses additional data on reservoir location and the avoided cost of sediment removal to values a sub-watershed's capacity to keep sediments out of reservoirs.

How it works

First, we estimate the potential for soil loss based on geomorphological and climate conditions. The model is based on the USLE, and represents the first four factors in the equation (rainfall erosivity, soil erodibility, and the length-slope factor). This part of the model accounts for two key relationships. In areas where rainfall intensity is high, there is a high chance that soil particles will become detached and transported by overland runoff. Also, in areas where the soil has a high proportion of sand, the erodibility is going to be low which means soil particles are easily detached from the soil pack and transported by overland runoff.

The Universal Soil Loss Equation (USLE) provides the foundation of the biophysical step of the InVEST sediment retention model.

$$USLE = R \times K \times LS \times C \times P \quad \text{(from Wischmeier \& Smith, 1978)}$$

where*R is the rainfall erosivity, K is the soil erodibility factor, LS is the slope length-gradient factor, C is the crop/vegetation and management factor and P is the support practice factor.*

The Slope Length Factor (LS) is one of the most critical parameters in the USLE. Slope length is the distance from the origin of overland flow along its flow path to the location of either concentrated flow or deposition. It reflects the indirect relationship between slope and land management (terracing, ditches, buffers, barriers). The LS factor is essentially the distance that a drop of rain/sediment runs until its energy dissipates. It represents a ratio of soil loss under given conditions compared to a reference site with the "standard" slope of 9% and slope length of 72.6 feet. The steeper and longer the slope is, relative to the conditions of the reference site, the higher the risk for erosion will be (for more information see http://www.omafra.gov.on.ca/english/engineer/facts/00-001.htm). The estimates of slope-length are based on methodology in a model called N-SPECT such that abrupt changes in slope result in length cutoffs. *Adjustments are necessary when slope is greater than 9% and slope length is different than 72.6 feet (22.12m).* In the model, different LS equations are automatically used for slope conditions that differ from the standard reference site conditions of the USLE equation development.

For low slopes: (The slope threshold should be specified as a model input and depends on the geomorphology and watershed characteristics)

$$LS = \left(\frac{flowacc * cellsize}{22.13}\right)^{nn} \left(\left(\frac{\sin(slope * 0.01745}{0.09}\right)^{1.4}\right) * 1.6$$

$$nn = \begin{cases} 0.5, slope >= 5\% \\ 0.4, 3.5 < slope < 5\% \\ 0.3, 1 < slope, = 3.5\% \\ 0.2, slope <= 1\% \end{cases}$$

where *flowacc* is accumulated water flow to each cell and *cellsize* is the pixel size or the grid resolution (10m, 30m, 90m, etc.) .

For high slopes: The slope threshold that the model uses to switch between the equation above and the following equation is specified as a model input and depends on the local geomorphology and watershed characteristics. We use the following equation, defined by Huang and Lu (1993) for areas with slopes higher than the threshold identified by the user:

$$LS = 0.08\lambda^{0.35} prct\_slope^{0.6}$$

$$\lambda = \begin{cases} cellsize, flowdir = 1,4,16 or 64 \\ 1.4 * cellsize, otherflowdir \end{cases}$$

where *prct_slope* is the pixel's percent slope and *flowdir* is the flow direction of the pixel

Calculation of Potential Soil Loss

We estimate the ability of vegetation to keep soil in place on a give parcel by comparing erosion rates on that parcel to what erosion rates would be on that parcel with no vegetation present (bare soil). The bare soil estimate is calculated as follows:

RKLS = R x K x LS

Erosion from the parcel with existing vegetation is calculated by the USLE equation:

USLE =R x K x LS x C x P

Avoided erosion (sediment retention) on the parcel is then calculated by subtracting USLE from RKLS.

Vegetation does not only keep sediment from eroding where it grows. It also traps sediment that has eroded upstream. The USLE equation overlooks this component of sediment dynamics, so we attempt to account for it as follows. All soil that the USLE equation estimates will erode is routed downstream via a flowpath. We estimate how much of the sediment eroded on all parcels will be trapped by downstream vegetation based on the ability of vegetation in each parcel to capture and retain sediment. The model also determines the total sediment load exported that reaches the stream from each pixel on the landscape. The table below describes how the removal of sediment by vegetation along hydrologic flowpaths is done:

Cell	Vegetation retention efficiency	USLE	Retained by Cell (retained)	Outflow quantity (OQ) from Cell (Gi=1-Ei)
1	E1	USLE1	0	USLE1
2	E2	USLE2	USLE1.E2	USLE1.G2+USLE2
3	E3	USLE3	((USLE1.G2+USLE2).E3	(USLE1.G2+USLE2).G3+USLE3
4	E4	USLE4	USLE1.G2.G3.E4+	USLE1.G2.G3.G4+

			USLE2.G3.E4+ USLE3.E4	USLE2.G3.G4+ USLE3.G4+ USLE4

The total retained sediment (tot_retain_x) is equal to the sum of the sediment removed by the pixel itself and the sediment removed through routing filtration.

The reservoir dead volume is designed to be filled completely at the end of the life time of the reservoir then the reservoir will loose its service in generating hydropower. We are assuming that each pixel on the landscape has the allowance of sediment equal to the dead volume divided by the time span of the reservoir and the number of cells in the reservoir watershed (max_allowance) to export the reservoir. So the portion of the sediment retention that the pixel will be valued at is the result of the subtraction of this max_allowance from the total retention.

The model then sums and averages the sediment export and retention per pixel to the sub-watersheds.

The valuation model uses the cost of sediment removal per m^3 entered by the user to determine the avoided cost per tons of sediment retained by the landscape.
The following equation is used to determine the value each pixel contributes to reservoir maintenance by helping to avoid erosion.

$$PVSR_x = \sum_{t'}^{T-1} \frac{SEDREM_x \times MC}{(1+r)^t}$$

where $PVSR_x$ is the present value of sediment retention on pixel x over T years, where T indicates the period of time over which the LULC pattern is constant or the length of the reservoir life length, $SEDREM_x$ is tot_retain minus max_allowance for pixel x, and MC is the marginal cost of sediment removal. This cost may vary across reservoirs in a single watershed if different technologies are employed for sediment removal at different reservoirs. If this is the case, the user may input reservoir-specific removal costs. The marginal cost of sediment removal should be measured in units of monetary currency per cubic meter (i.e. $\$/m^3$). The discount rate is r.

The model generates the retention value per sub-watershed by summing the value per pixel for all pixels within each of the sub-watersheds.

Limitations and simplifications

Although the USLE method is a standard way to calculate soil loss, it has several limitations. The USLE method predicts erosion from sheet wash alone (erosion from plains in gentle slopes) (FAO 2002). Rill-inter-rill, gullies and/or stream-bank erosion/deposition processes are not included in this model. As such, it is more applicable to flatter areas because it has only been verified in areas with slopes of 1 to 20 percent. Moreover, the relationship between rainfall intensity and kinetic energy may not hold in mountainous areas because it has only been tested in the American Great Plains. Finally, the equation considers only the individual effect of each variable. In reality, some factors interact with each other, altering erosion rates.

Another simplification of the model is the grouping of LULC classes because the model's results are highly sensitive to the categorization of LULC classes. If there is a difference in land use between two areas within the same broad LULC category, it is recommended to create two LULC categories. For example, if all forest is combined into one LULC class, the difference in soil retention between an old growth forest and a newly planted forest is neglected. More generally, where there is variation across the landscape that affects a USLE parameter, the LULC classes should reflect that variation.

Third, the model relies on retention or filtration efficiency values for each LULC type. However, there are often few data available locally for filtration rates associated with local LULC types. Data from other regions may be applied in these cases, but may misrepresent filtration by local LULC types.

Additionally, the model may not accurately depict the sedimentation process in the watershed of interest since the model is based on parameterization of several different equations and each parameter describes a stochastic process. Due to the uncertainty inherent in the processes being modeled, it is not recommended to make large-scale area decisions based on a single run of the model. Rather, the model functions best as an indicator of how land use changes may affect the cost of sediment removal, and like any model is only as accurate as the available input data. A more extensive study may be required for managers to calculate a detailed cost-benefit analysis for each reservoir site.

Another assumption is that sediment retention upstream from a reservoir is valuable only if sediment delivery impacts reservoir function, which incurs a cost. If sediment is not removed from a reservoir, the model does not assign a value to the sediment retention service. In this case, the user may assign a value to upstream sediment retention based on an assumed trajectory of sediment deposition at the reservoir. This method is explained below and it not included in this model. As noted above, we are only modeling sheetwash erosion, meaning that our estimate of annual reservoir sedimentation will be less than actual sedimentation rates. Nonetheless, it is possible to use information about the sediment volume in the reservoir at time t, V_t , and the volume at which reservoir function will be impacted, V_D , to estimate the time period over which sediment removal will occur. If the user is able to provide accurate estimates of V_t and V_D, then it is likely that information about annual deposition is available as well. Let $SEDDEP_t$ represent the total volume of sediment (USLE) assumed to reach the reservoir in a given year. Then we can model the time path of sediment as $V_{t+1} = SEDDEP_t + V_t$, and we can define the year at which removal will commence, *t'*, as the first period for which $V_t > V_D$. In this case, let the value of sediment retention on the upstream parcel *x* be given by $PVSR_{x \in d} = \sum_{t=t'}^{T-1} \frac{SEDREM_{jx} \times MC_d}{(1+r)^t}$

where, $PVSR_x$ is the present value of sediment retention on pixel *x* over *T* years, where *T* indicates the period of time over which the LULC pattern is constant or the length of the reservoir life length. $SEDREM_x$ is the sediment removed by the LULC on pixel *x*. *MC* is the marginal cost of sediment removal. *r* is the discount rate.

The accuracy of the sediment retention value is limited by two factors. First, it is limited by the quality of information of the cost of sediment removal. Up-to-date estimates of sediment removal costs for an area may be difficult to find. The user may be limited to using an outdated average value from other locations and for a different type of reservoir. Second, the accuracy of the model is limited by the user's ability to calibrate it with actual sedimentation data. The model allows for a calibration constant to be applied and adjusted via the Sediment Delivered output. This can greatly improve the model, but only if the user has access to reliable sedimentation data for the watershed(s) of interest.

Data needs

There are nine required data inputs for this model and one optional input (required for valuation). 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. 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 necessary 'burning' hydrographic features into the elevation model (recommended when you see unusual streams.) See the Working with the DEM section of this manual for more information.

Name: File can be named anything, but no spaces in the name and less than 13 characters
Format: standard GIS raster file (e.g., ESRI GRID or IMG), with elevation value for each cell given in meters above sea level.
Sample data set: C:\Invest\Base_Data**dem**

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. See the Appendix for further details.

Name: File can be named anything, but no spaces in the name and less than 13 characters Format: standard GIS raster file (e.g., ESRI GRID or IMG), with a rainfall erosivity index value for each cell given in $MJ*mm*(ha*h*yr)^{-1}$.
Sample data set: C: \Invest\Sedimentation\Input**erosivity**

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.

> Name: File can be named anything, but no spaces in the name and less than 13 characters
> Format: standard GIS raster file (e.g., ESRI GRID or IMG), with a soil erodibility value for each cell. K is in $T.ha.h.\ (ha.MJ.mm)^{-1}$.
> **Sample data set**: C:\Invest\Sedimentation\Input**erodibility**

4. Land use/land cover (LULC) (required). LULC is a GIS raster dataset, with an integer LULC code for each cell.

Name: File can be named anything, but no spaces in the name and less than 13 characters
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 Biophysical table (see below).
Sample data set: C:\Invest\Base_Data**landuse_90**

5. Sub-watersheds (required). This is a shapefile that contains sub-watersheds for each point of interest for the sediment valuation model. These points represent reservoirs or other types of structures where sedimentation is an issue of concern. The sub-watersheds are delineated within each watershed contributing to these points of interest.

Due to limitations in ArcMap geoprocessing, the maximum size of a watershed or sub-watershed to be processed in this model is approximately the equivalent of 4000x4000 cells, with cell size equal to the smallest cell size of your input layers.. If the whole watershed contributing to a point of interest is larger than this size, it will need to be divided into sub-watersheds that are each smaller. Then, the resulting sub-watershed layer should be entered here and the whole watershed layer should be used in the Watersheds input. If the whole watershed is smaller, then it does not need to be divided, and the same watershed layer should be entered for both Sub-watersheds and Watersheds inputs. Sub-watersheds will be mosaiced back together into whole watersheds for the final output.

Name: File can be named anything, but no spaces in the name.
Format: shapefile of polygons where the subws_id field contains a unique integer ID value for each sub-watershed.
Sample data set: C:\Invest\Base_Data**subwatersheds.shp**

6. Watersheds (required). A shapefile that contains one watershed for each point of interest for the sediment valuation model. These points represent reservoirs or other types of structures where sedimentation is an issue of concern. The watersheds are delineated for these points of interest. See **Sub-watersheds** above for more information on watershed size requirements.

Name: File can be named anything, but no spaces in the name .
Format: shapefile of polygons where the ws_id field contains a unique integer ID value for each watershed, which corresponds to the watershed/reservoir ID field in the Sediment Valuation input table.
Sample data set: C:\Invest\Base_Data**watersheds.shp**

7. Biophysical table (required). A table containing model information corresponding to each of the land use classes. NOTE: these data are attributes of each LULC class, not each cell in the raster map.

Name: Table names should only have letters, numbers and underscores, no spaces
File type: *.dbf or *.mdb

Rows: Each row is a land use/land cover class.
Columns: Each column contains a different attribute of each land use/land cover class and must be named 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 above.
b. *LULC_desc*: Descriptive name of land use/land cover class (optional)
c. *usle_c:* Cover and management factor for the USLE. This value is given in the table as an integer by multiplying the C factor by 1000.
d. *usle_p:* Management practice factor for the USLE. This value is given in the table as an integer by multiplying the P factor by 1000.
c. *sedret_eff:* The sediment retention value for each LULC class, as an integer percent between zero and 100. This field identifies the capacity of vegetation to retain sediment, as a percentage of the amount of sediment flowing into a cell from upslope. In the simplest case, when data for each LULC type are not available, a value of 100 may be assigned to all natural vegetation types (such as forests, natural pastures, wetlands, or prairie), indicating that 100% of sediment is retained.

An intermediary value also may be assigned to features such as contour buffers. All LULC classes that have no filtering capacity, such as pavement, can be assigned a value of zero.

Sample data set: C: \Invest\Base_Data\Water_Tables.mdb**Biophysical_Models**

8. Threshold flow accumulation (required). The number of upstream cells that must flow into a cell before it's considered part of a stream. Used to define streams from the DEM. The model's default value is 1000. This value is to define and create the stream lines. If the user has a map of streams in the watershed of interest, he/she should compare it the V_stream map (output of the model). This value also needs to be well estimated in watersheds where ditches are present. This threshold expresses where hydraulic routing is discontinued and where retention stops and the remaining pollutant will be exported to the stream.

9. Slope threshold (required). An integer slope value describing landscape characteristics such as slope management practices including terracing and slope stabilization techniques. It depends on the DEM resolution and the terracing practices used in the region. In so many cases around the world, farmers cultivate slopes without any terracing or slope stabilization but after certain slope limit, terraces become necessary. This limit could be taken as this slope threshold. A good understanding, knowledge and familiarity with the regional landscape will help the user to determine this parameter. The model's default value is 25. This threshold could a subject for calibration.

10. Sediment valuation table A table containing valuation information for each of the reservoirs. There must be one row for each watershed in the Watersheds layer.

Name: Table names should only have letters, numbers and underscores, no spaces
File type: *.dbf or *.mdb
Rows: Each row is a reservoir or structure that corresponds to the watersheds raster.
Columns:

a. *ws_id* (watershed ID): Unique integer value for each reservoir, which must correspond to values in the Watersheds layer.
b. *dr_cost*:: Cost of sediment dredging in $ (Currency) / m³ removed. Floating point value. Used for valuing sediment retention for dredging.
c. *dr_deadvol:* The volume of water below the turbine. It is a design dimension below which water is not available for any use and it's designed to store (deposit) sediment without hindering turbine and reservoir hydropower functions. Used for valuing sediment retention for dredging.
d. *dr_time*: Integer time period to be used in calculating Present Value (PV) of removal costs. This time period should be the remaining designed lifetime of the structure. For instance, if you are using an LULC map for the year 2000 and a reservoir of interest was designed in 1950 for a 100-year lifetime, the time period entered here should be 50 years. Used for valuing sediment retention for dredging.
e. *dr_disc:* The rate of discount over the time span, used in net present value calculations. Used for valuing sediment retention for dredging. Floating point value.
f. *wq_cost*:: Cost of sediment dredging in $ (Currency) / m³ removed. Floating point value. Used for valuing sediment retention for water quality.
g. *wq_annload:* Used for valuing sediment retention for water quality.
h. *wq_time*: Integer time period to be used in calculating Present Value (PV) of removal costs. This time period should be the remaining designed lifetime of the structure. For instance, if you are using an LULC map for the year 2000 and a reservoir of interest was designed in 1950 for a 100-year lifetime, the time period entered here should be 50 years. Used for valuing sediment retention for dredging. Used for valuing sediment retention for water quality.
i. *wq_disc:* The rate of discount over the time span, used in net present value calculations. Used for valuing sediment retention for water quality. Floating point value.

<u>**Sample data set:**</u> C: \Invest\Base_Data\Water_Tables.mdb**Sediment_Valuation**

11. Sediment threshold table A table containing annual sediment load threshold information for each of the reservoirs. There must be one row for each watershed in the Watersheds layer.

<u>Name</u>: Table names should only have letters, numbers and underscores, no spaces
<u>File type:</u> .dbf or .mdb
<u>Rows</u>: Each row is a reservoir or structure that corresponds to the watersheds layer.
<u>Columns</u>:

a. *ws_id* (watershed ID): Unique integer value for each reservoir, which must correspond to values in the Watersheds layer.
b. *dr_time*: Integer time period corresponding to the remaining designed lifetime of the structure. For instance, if you are using an LULC map for the year 2000 and a reservoir of interest was designed in 1950 for a 100-year lifetime, the time period entered here should be 50 years.
c. *dr_deadvol:* The volume of water below the turbine. It is a design dimension below which water is not available for any use and it's designed to store (deposit) sediment

without hindering turbine and reservoir hydropower functions. Used for valuing sediment retention for dredging.

<u>Sample data set:</u> C: \Invest\Base_Data\Water_Tables.mdb**Sediment_Threshold**

Running the Model

The Avoided Reservoir Sedimentation model maps the soil loss, sediment exported, sediment retained, and value of sediment retention on the landscape. This model is structured as a toolkit which has two tools. The first tool, Soil Loss, produces multiple outputs, including USLE, sediment retained by the landscape and sediment exported to the stream. One of these output values feed into the next portion of the model, the Valuation tool, which calculates sediment retention value. By running the tool, three folders will automatically be created in your workspace (you will have the opportunity to define this file path): "Intermediate", where temporary files are written and which is deleted after each tool run; "Service", where results that show ecosystem services are saved (such as sediment retention); and "Output", where non-service biophysical results are saved (such as sediment export.)

Before running the Avoided Reservoir Sedimentation Model, make sure that the InVEST toolbox has been added to your ArcMap document, as described in the **Getting Started** chapter of this manual. Second, make sure that you have prepared the required input data files according to the specifications in *Data Needs*.

- Identify workspace

  If you are using your own data, you need to first create a workspace, or folder for the analysis data, on your computer hard-drive. The entire pathname to the workspace should not have any spaces. All your output files will be saved here. For simplicity, you may wish to call the folder for your workspace 'Sediment' and create a folder in your workspace called "Input" and place all your input files here. It's not necessary to place input files in the workspace, but advisable so you can easily see the data you use to run your model.

  Or, if this is your first time using the tool and you wish to use sample data, you can use the data provided in InVEST-Setup.exe. If you unzipped the InVEST files to your C-drive (as described in the **Getting Started** chapter), you should see a folder called /Invest/**Sediment**. This folder will be your workspace. The input files are in a folder called /Invest/Sediment/**Input** and in /Invest/Base_Data/Freshwater.

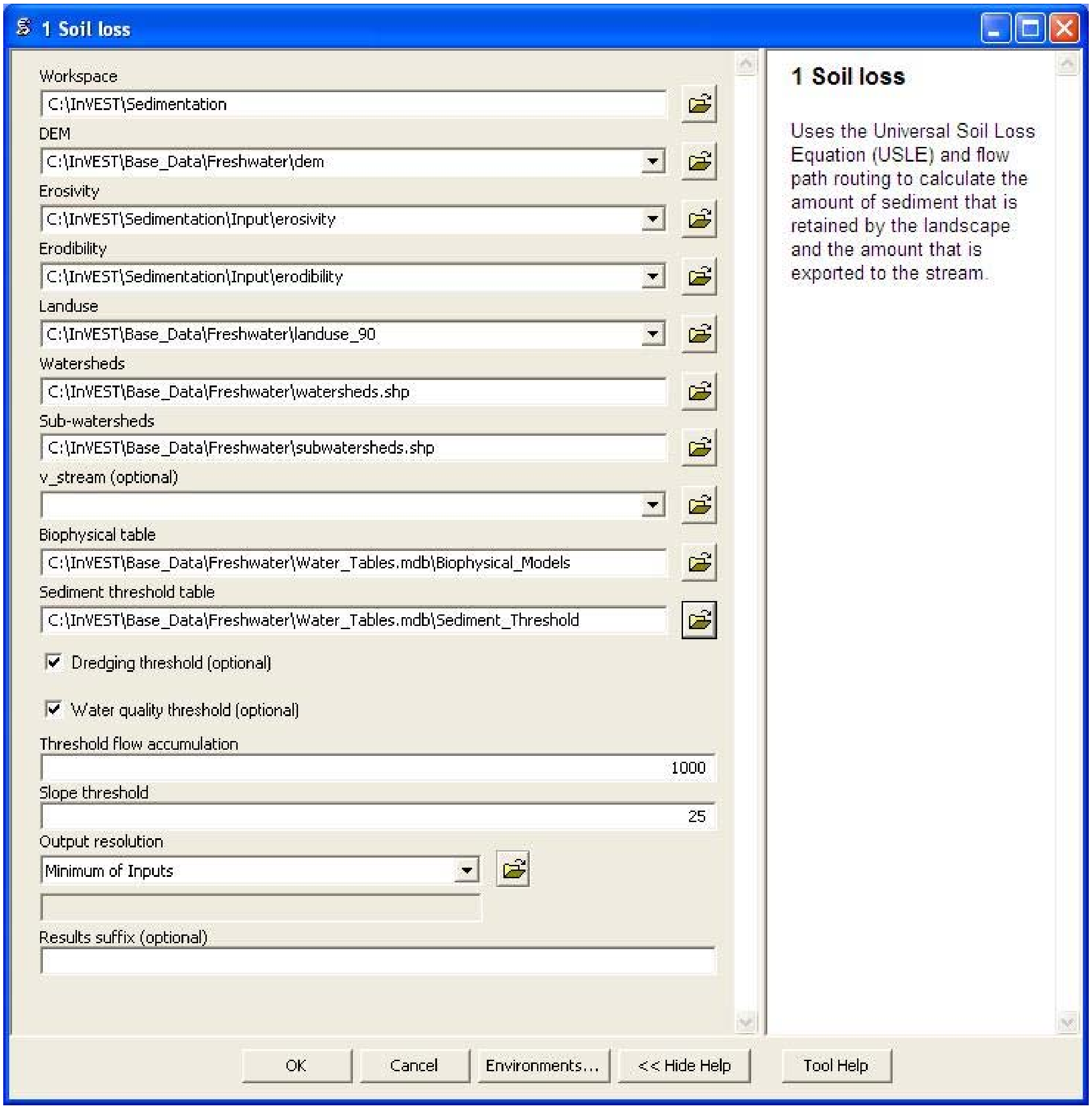
- Open an ArcMap document to run your model.

- Find the InVEST toolbox in ARCToolbox. ARCToolbox is normally open in ARCMap, but if it is not, click on the ARCToolbox symbol. See the **Getting Started** chapter if you don't see the InVEST toolbox and need instructions on how to add it.

- You can run this analysis without adding data to your map view, but usually it is recommended to view your data first and get to know them. Add the data for this analysis to your map using the ADD DATA button and look at each file to make sure it is formatted correctly. Save your ARCMap file as needed.

- Click once on the plus sign on the left side of the INVEST toolbox to see the list of tools expand. Next, click on the plus sign next to the Avoided Reservoir Sedimentation toolset. Within the toolset are two tools, Soil Loss and Valuation. You will need to run Soil Loss first to generate layers that will feed into Valuation.

- Double click on Soil Loss. An interface will pop up like the one below. The tool shows default file names, but you can use the file buttons to browse instead to your own data. When you place your cursor in each space, you can read a description of the data requirements in the right side of the interface. **Click Show Help if the description isn't showing by default.** In addition, refer to the ***Data Needs*** section above for information on data formats.

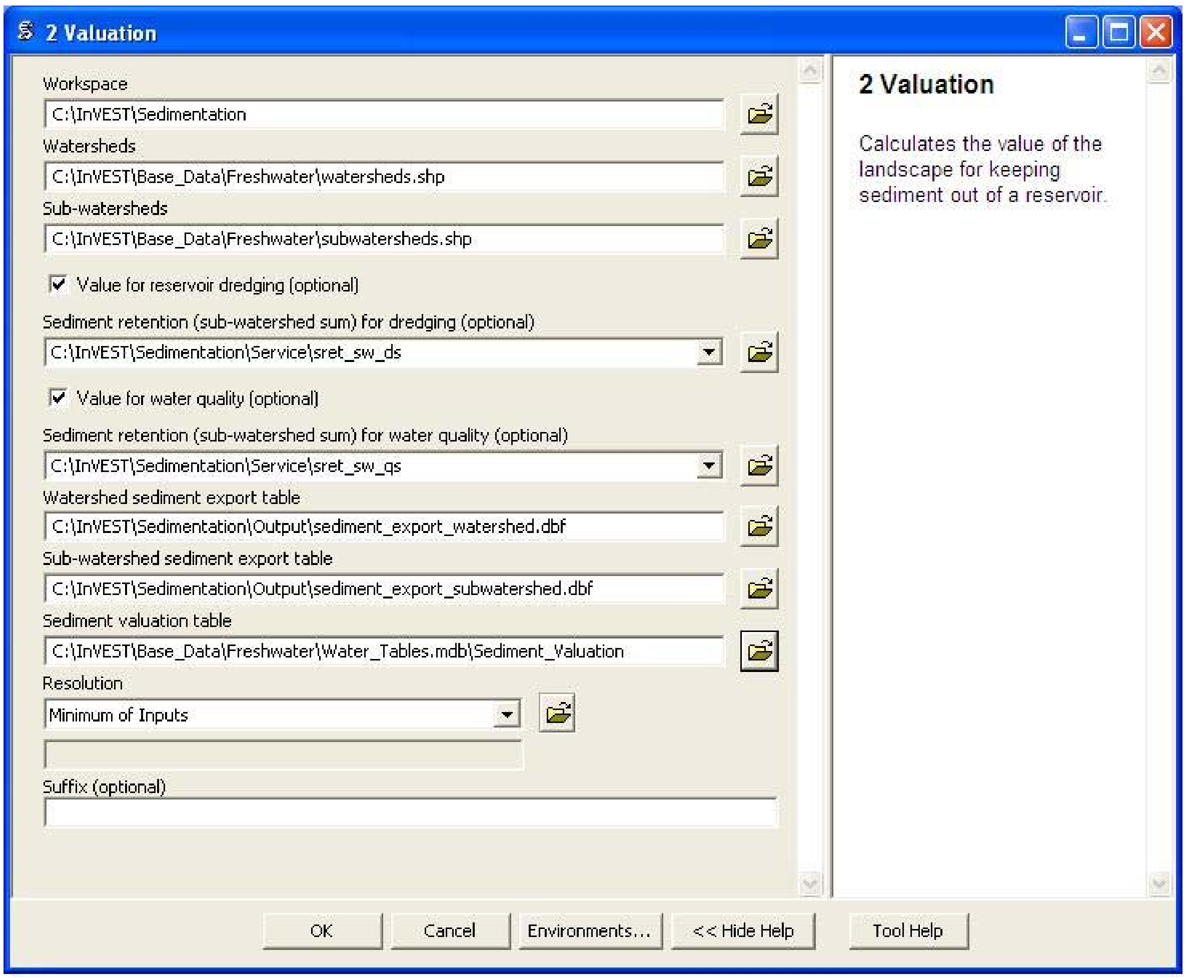


- Fill in data file names and values for all required prompts. Unless the space is indicated as optional, it requires you to enter some data.
- After you've entered all values as required, click on OK. The script will run, and its progress will be indicated by a "Progress dialogue".
- Upon successful completion of the model, you will see new folders in your workspace called "Intermediate", "Service" and "Output". These folders contain several raster grids. These grids are described in the next section.
- Load the output grids into 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 file appears in the map.

- You can also view the attribute data of output files by right clicking on a layer and selecting OPEN ATTRIBUTE TABLE.

- Now, run the Valuation Tool. Either three or four outputs from the Soil Loss model are inputs to this model, depending on whether dredging, water quality or both are valued: sret_sw_qs (sediment retention for water quality, summed by sub-watershed), sret_sw_ds (sediment retention for dredging, summed by sub-watershed), sediment_export_watershed.dbf (table of sediment export per watershed) and sediment_export_subwatershed.dbf (table of sediment exort per sub-watershed.) . The interface is below:



- When the script completes running, the outputs will be placed into the “Service” folder. A description of the files is below.

- Since this model is open source, the user can edit the scripts to modify, update, and/or change equations by right clicking on the script's name and selecting "Edit…" The script will then open in a text editor. After making changes, click File/Save to save your new script.

Interpreting Results

The following is a short description of each of the outputs from the Avoided Reservoir Sedimentation model, all of which are automatically saved into the "Outputs" folder in your workspace:

Output\ **usle_sws_m** (tons/ha): Mean potential soil loss per sub-watershed.

Output**usle_sws_s**: (tons): Total potential soil loss per sub-watershed.

Output**sediment_export_watershed.dbf:** Table containing the total sediment exported (sed_load) to the outlet of each watershed. This sed_load will 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.

Output**sediment_export_subwatershed.dbf:** Table containing the total sediment exported (sed_load) within each sub-watershed.Output**upret_sw_m** (tons/ha): Raster containing the mean amount of sediment retained from sediment originating upstream of each pixel, averaged across pixels in each sub-watershed. Does not include the sediment originating from the pixel itself.

Output**upret_sw_s** (tons): Raster containing the total amount of sediment retained from sediment originating upstream of each pixel, summed across pixels in each sub-watershed. Does not include the sediment originating from the pixel itself.

Output**sret_sw_qm** (Sediment Retained) (tons/ha): Raster containing the mean sediment retained on each sub-watershed, including sediment retained that originates upstream as well as sediment that originates on the cell itself. It is adjusted by the water quality sediment allowable threshold.

Output**sret_sw_qs** (Sediment Retained) (tons/sub-watershed): Raster containing the total sediment retained within each sub-watershed, including sediment retained that originates upstream as well as sediment that originates on the cell itself. It is adjusted by the water quality sediment allowable threshold.

Output**sret_sw_dm** (Sediment Retained) (tons/sub-watershed): Raster containing the mean sediment retained per cell on each sub-watershed, including sediment retained that originates upstream as well as sediment that originates on the cell itself. It is adjusted by the reservoir dead volume allowance.

Output**sret_sw_ds** (Sediment Retained) (tons/sub-watershed): Raster containing the total sediment retained within each sub-watershed, including sediment retained that originates upstream as well as sediment that originates on the cell itself. It is adjusted by the reservoir dead volume allowance.

Output\ **sexp_sw_m** (tons/ha): Raster containing the mean sediment export per cell on each sub-watershed.

Output\ **sexp_sw_s** (tons): Raster containing the total sediment export within each sub-watershed.

Service**sed_val_dr** (Value of Sediment Removal for dredging): Raster showing the value ($(Currency) per sub-watershed) of the landscape for retaining sediment by keeping it from entering the reservoir, thus avoiding dredging costs. **THIS IS THE SUB-WATERSHED MEASURE OF THIS ECOSYSTEM SERVICE IN ECONOMIC TERMS.**

Service**sed_val_wq** (Value of Sediment Removal for water quality sediment standard): Raster showing the value ($(Currency) per sub-watershed) of the landscape for retaining sediment by keeping it from entering the reservoir, thus avoiding water treatment costs. **THIS IS THE SUB-WATERSHED MEASURE OF THIS ECOSYSTEM SERVICE IN ECONOMIC TERMS.**

Service**sediment_value_watershed.dbf:** Table of sediment values for each watershed: total export, total retention and total retention value.

Service**sediment_value_subwatershed.dbf:** Table of sediment values for each sub-watershed: total export, total retention and total retention value.

The application of these results depends entirely on the objective of the modeling effort. Users may be interested in all of these results or select one or two. If sediment removal cost information is not available or valuation is not of interest, the user may use a value of one for the cost of sediment removal. This forces a unit cost of sediment removal, which normalizes the cost across the different reservoirs but still allows a relative comparison of scenarios.

The following provides more detail on each of the relevant model outputs. The length-slope factor depends solely on the geometry of the landscape, and, as the name infers, is simply a description of the length of the slopes in the watershed. The RKLS is the potential soil loss based on the length-slope factor, rainfall erosivity, and soil erodibility. These are factors that generally cannot be altered by human activity, as they are inherent to the watershed.

USLE differs from RKLS in that it takes into account the management practice factor and the cover factor. These are factors that can be altered with land use changes or changes in land management. Examples of changes that can alter the USLE output are forest clear cuts, changing crop type or type of agriculture (no till to tilled), expansion of an urban area, or restoring vegetation along a stream-bank. The model output describes this 'actual' soil loss on an annual basis in tons per hectare, summarized in a raster grid over the landscape.

The user should understand that this USLE method predicts the sediment from sheet wash alone. Rill-inter-rill, gullies and/or stream-bank erosion/deposition processes are not included in this model. A visit to the watershed and consultation of regional research results need to be used to evaluate the portion of sheet wash in the total sediment loading that is used in testing and verifying this model.

Total Sediment exported to the outlet of the watershed (sediment delivered) indicates the volume of soil delivered each year. Since this model doesn't simulate the in-stream processes where erosion and deposition could have a major impact on this sediment delivered, the user should pay great attention to their importance while calibrating or adjusting this model. When soil deposition rates are known from observations at interest points, the user can aggregate the Sediment Delivered values (tons of sediment) and compare to observations. Remember that USLE only predicts sheet erosion (not landslide or roads induces or channel erosion), so a sediment budget (distribution of observed sediment yield into erosion types) must be performed to compare the correct measured sources of sediment with the model output.

The Value of Sediment Removal is a raster grid that displays the present value (in currency per pixel) of sediment retention on the landscape. In other words, it is the avoided cost of sediment removal at a downstream reservoir (over the reservoirs projected lifetime) due to the ability of the landscape to keep sediment in place. This raster grid provides valuable information to the decision maker on the relative importance of each part of the landscape in determining the cost of sediment removal for a particular reservoir. This output allows managers to see which parts of the landscape are providing the greatest value in terms of avoided sediment removal costs. They may want to protect, or at least avoid serious land use change, in these areas. Similarly, when scenarios of future land management are analyzed with this model, the Value of Sediment Removal layer can be used to identify where the benefits of avoided maintenance costs will be lost, maintained or improved across the landscape. Summarizing this layer across the landscape can also give an overall sense of the total costs that will be avoided given a particular landscape configuration.

The user should keep in mind that the Tier 1 model may not accurately depict the sedimentation process in the user's watershed of interest. Furthermore, the model is based on parameterization of several different equations, and each parameter describes a stochastic process. Due to the uncertainty inherent in the processes being modeled here, the user should not make large-scale decisions based on a single run of this model. The Avoided Reservoir Sedimentation model provides a first cut in prioritization and comparison of landscape management alternatives. A more detailed study is required for managers to calculate a specific benefit-cost analysis for each reservoir site. This model functions best as an indicator of how land use changes may affect the cost of sediment removal, and like any model is only as accurate as the available input data.

Appendix: 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.

1. 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 the World Wildlife Fund - http://www.worldwildlife.org/freshwater/hydrosheds.cfm. NASA provides free global 30m DEM data at http://asterweb.jpl.nasa.gov/gdem-wist.asp. Or, it may be purchased relatively inexpensively at sites such as MapMart (www.mapmart.com). The DEM resolution is 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.

2. Rainfall erosivity index (R)

R should be obtained from published values, as calculation is very tedious. For calculation, R equals E (the kinetic energy of rainfall) times I30 (maximum intensity of rain in 30 minutes in cm/hr). Roose (1996) found that for Western Africa R = a * precipitation where a = 0.5 in most cases, 0.6 near the sea, 0.3 to 0.2 in tropical mountain areas, and 0.1 in Mediterranean mountain areas.

The following equation is widely used to calculate the R index (http://www.fao.org/docrep/t1765e/t1765e0e.htm):

$$R = E*I30 = (210 + 89 \log_{10}I30)*I30$$

E: kinetic energy of rainfall expressed in metric MJ× m/ha/cm of rainfall.
I30: maximum intensity of rain in 30 minutes expressed in cm per hour.

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 (multiply by 17.02). 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.

3. 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. It is 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. Values of 0 – 0.6 are reasonable, while higher values should be given a critical look. K may be found as part of standard soil data maps.

Coarse, yet free global soil characteristic data is available at http://www.ngdc.noaa.gov/seg/cdroms/reynolds/reynolds/reynolds.htm. The FAO also provides global soil data in their Harmonized World Soil Database: http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/ .
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://soils.usda.gov/survey/geography/ssurgo/ and STATSGO http://soils.usda.gov/survey/geography/statsgo/ . 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 erodibility should be calculated as the average of all horizons within a soil class component, and then a weighted average of the components should be estimated. This can be a tricky GIS analysis: In the US soil categories, each soil property polygon can contain a number of soil type components with unique properties, and each component may have different soil horizon layers, also with unique properties. Processing requires careful weighting across components and horizons. The Soil Data Viewer (http://soildataviewer.nrcs.usda.gov/), a free ArcMap extension from the NRCS, does this soil data processing for the user and should be used whenever possible.

The following equation can be used to calculate K (Wischmeier and Smith 1978):

$$K = 27.66 * m\text{^}1.14 * 10\text{^}{-8} * (12 - a)) + (0.0043 * (b - 2)) + (0.0033 * (c - 3))$$

In which
K = soil erodibility factor (t*ha/MJ*mm)
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.

4. Land use/land cover

A key component for all 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 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.umiacs.umd.edu/data/landcover/. This data is available in 1 degree, 8km and 1km resolutions. 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 could involve 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. The user should only break up a land use type if it will provide more accuracy in modeling. For instance, for the sediment valuation 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.

5. P and C coefficients

The management 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 and 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:

• U.S. Department of Agriculture soil erosion handbook
http://topsoil.nserl.purdue.edu/usle/AH_537.pdf

• USLE Fact Sheet
http://www.omafra.gov.on.ca/english/engineer/facts/00-001.htm

• U.N. Food and Agriculture Organization
http://www.fao.org/docrep/T1765E/t1765e0c.htm

The final PC coefficient used in the model should be P x C x 1000, to ensure integer values.

6. Vegetation retention efficiencies

These values are used to incorporate the effects of natural vegetation that buffer potential water quality impairment downhill from sources. To develop these values, all land class pixels that contain natural vegetation (such as forests, natural pastures, wetlands, or prairie) are assigned high values and vegetation that has no or little filtering value receives a value of zero. All values should fall between 0 and 100. Consult with a hydrologist if not certain about assignment of specific values.

7. Watersheds

The watersheds should be delineated by the user, based on the location of reservoirs or other points of interest. 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://crunch.tec.army.mil/nidpublic/webpages/nid.cfm).

Watersheds that contribute to the points of interest must be generated. If known correct watershed maps exist, they should be used. Otherwise, watersheds can be generated in ArcMap using a hydrologically-correct digital elevation model. Due to limitations in ArcMap geoprocessing, the maximum size of a watershed that can be processed by the Nutrient Retention tool is approximately the equivalent of 4000x4000 cells. If the whole watershed contributing to a point of interest is larger than this size, it will need to be divided into sub-watersheds that are each smaller. If the whole watershed is smaller, then it does not need to be divided. Sub-watersheds will be mosaicked back together into whole watersheds for the final output.

See the Working with the DEM section of this manual for more information on generating watersheds and sub-watersheds.

8. Sediment table

The estimated sediment removal cost from the reservoirs will ideally be based on the characteristics of each reservoir and regional cost data. The user should consult managers at the individual reservoirs or a local sediment removal expert. The technology available at each location may vary, and the applicability of the specific technologies depends on the storage capacity/mean annual runoff ratio and the storage capacity/Annual Sediment yield ratio.

Once a range of possible technologies has been established for each reservoir, the model user should investigate past sediment removal projects to determine appropriate costing. This may require calculating to present day value and taking into account that the technology may have improved, reducing the relative cost.

If local information is not available, pricing must be estimated using published information. Adjust costs to specific requirements, location, and present day value as needed.

References

Anderson, J.R., Hardy E., Roach, J., and Witmer, R. 1976. A Land Use and Land Cover Classification System For Use with Remote Sensor Data: Geological Survey Professional Paper 964. Edited by NJDEP, OIRM, BGIA, 1998, 2000, 2001, 2002, 2005

FAO. 2002., *FAOSTAT* Homepage of Food and Agriculture Organization of the United Nations, Online 2008, 9/11 .

Huang Yanhe and Lu Chenglong. 1993. Advances in the application of the Universal Soil Loss Equation (USLE) in China. Journal of Fujian Agricultural College

(Natural Science Edition) 22 (1): 73 ~ 77.
Roose, E. 1996, *Land Husbandry -Components and strategy. 70 FAO Soils Bulletin,* Food & Agriculture Organization of the UN, Rome, Italy.
Wischmeier, W.H. & Smith, D. 1978, Predicting rainfall erosion losses: a guide to conservation planning. USDA-ARS Agriculture Handbook , Washington DC.

MANAGED TIMBER PRODUCTION MODEL

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 valuation 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.

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.

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 polygon attribute table;

see below), frequency of harvest (*Freq_harv*), and harvest and management (or maintenance) costs (*Harv_cost* and *Maint_cost*, respectively) (Fig. 1).

The timber parcel map can either be associated with a current (sometimes referred to as "base") LULC map used in most other InVEST models (where the year associated with the current LULC

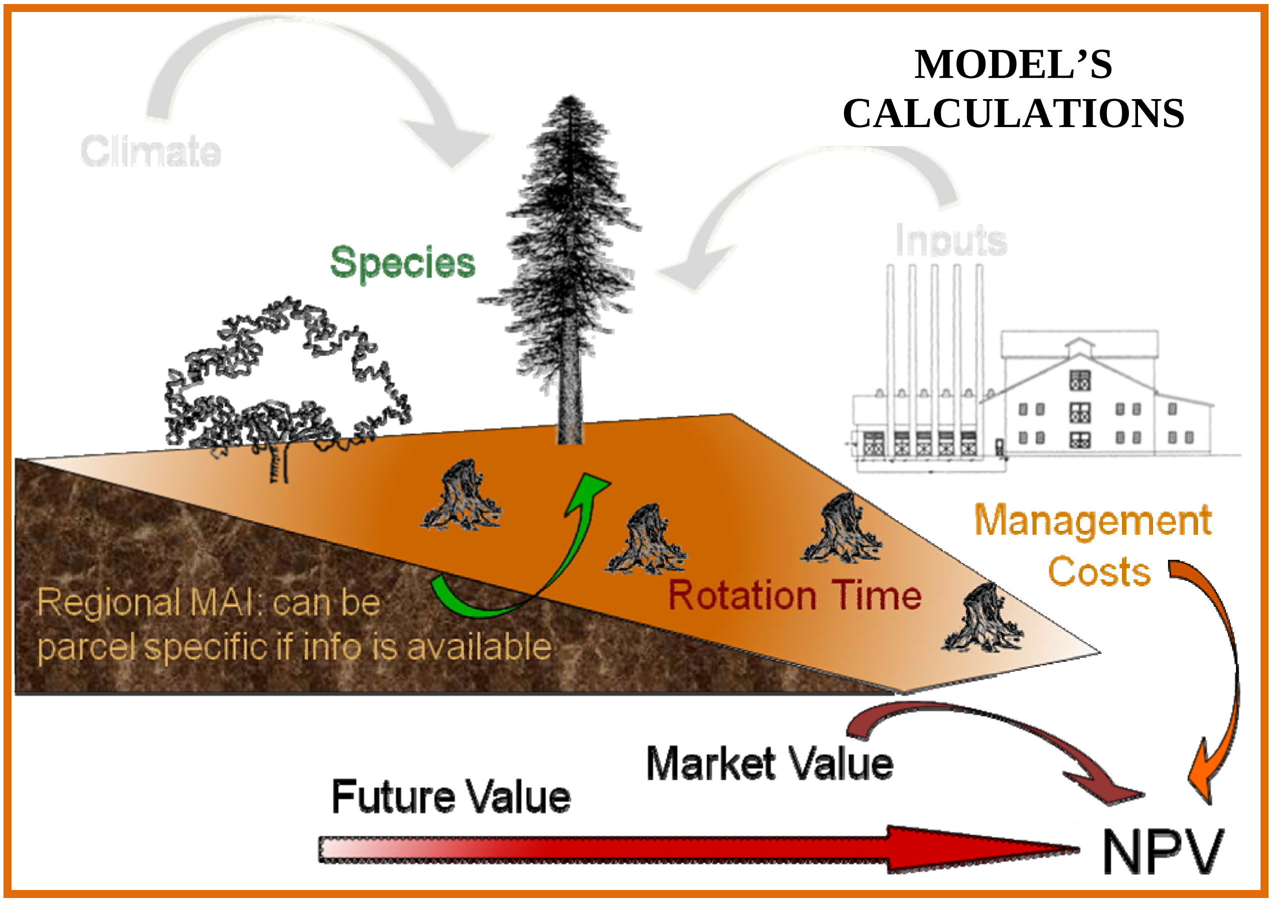


Figure 1. Conceptual diagram of the Managed Timber Production model. Parameters represented in color are included in the model, while those in gray are not.

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 modeledl. 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.

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, can be included as a table in the shapefile or as a separate 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. The dataset should be projected in meters and the projection used should be defined.

   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. The production table data containing attributes of the parcel can be included as part of the shapefile's attribute table or as a separate table that is joined or related to the shapefile. Either way, the variables and parameters to include in the data table are described below.

Sample data set: \Invest\Timber\Input**plantation.shp**

2. **Production table (required)**. A table of information about the timber parcels on the landscape. This is a separate data table that can be joined to the polygon dataset in #1.

   Name: file can be named anything

   File type: *.dbf, or an attribute table as part of the timber parcel map.

   Rows: each row is a different parcel.

   Columns: contain an attribute for each parcel and must be named as follows:

   a. *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.
   b. *Parcl_area*: The area of the timber parcel in hectares.
   c. *Perc_harv*: The proportion of the timber parcel area that is harvested each harvest period; units are integer percent.
   d. *Harv_mass*: The mass of wood harvested per hectare (in metric tons (Mg) ha^{-1}) in each harvest period.
   e. *Freq_harv*: The frequency of harvest periods, in years, for each parcel.
   f. *Price*: The marketplace value of the wood harvested from the parcel (Mg^{-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).
   g. *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.)
   h. *Harv_cost*: The cost (ha^{-1}) incurred when harvesting *Harv_mass*.
   i. *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).

j. *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*.

k. *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 m^3 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.dbf**

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*,

$$VH_x = \frac{Perc\_harv_x}{100}\left(Price_x \times Harv\_mass_x - Harv\_cost_x\right) \tag{1}$$

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), $Price_x$ 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 $Harv\_cost_x$ 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, $Price_x$ is expected to be \$10 Mg^{-1}, $Harv\_mass_x$ is 800 Mg ha^{-1}, and $Harv\_cost_x$ = \$5,000 ha^{-1}. The net value created during a harvest period is given by,

$$VH_x = 0.1 \times (10 \times 800 - 5000) = 300. \tag{2}$$

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 (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 VH_x intermittingly from yr_cur or yr_fut to T_x years after yr_cur or yr_fut is given by:

$$NPV_x = \sum_{s=0}^{\mathrm{ru}\left(\frac{T_x}{Freq\_harv_x}\right)-1} \frac{VH_x}{\left(1+\frac{r}{100}\right)^{(Freq\_harv_x \times s)}} - \sum_{t=0}^{T_x-1}\left(\frac{Mait\_cost_x}{\left(1+\frac{r}{100}\right)^t}\right) \qquad (3)$$

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 $Maint\_cost_x$ is the annualized cost (ha^{-1}) of managing parcel x. Continuing our earlier example, where 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 $Maint\_cost_x$ equal to \$50 ha^{-1}, then the NPV of the stream of VH_x is,

$$NPV_x = \sum_{s=0}^{9} \frac{300}{1.07^s} - \sum_{t=0}^{9} \frac{50}{1.07^t} \qquad (4)$$

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,

$$\begin{aligned} NPV_x &= \sum_{s=0}^{\mathrm{ru}\left(\frac{10}{3}\right)-1} \frac{300}{1.07^{3\times s}} - \sum_{t=0}^{9}\left(\frac{50}{1.07^t}\right) \\ &= \sum_{s=0}^{3} \frac{300}{1.07^{3\times s}} - \sum_{t=0}^{9}\left(\frac{50}{1.07^t}\right) \end{aligned} \qquad (5)$$

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 $\times$ $Freq_x$ 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,

$$NPV_x = \sum_{s=1}^{\mathrm{rd}\left(\frac{T_x}{Freq\_harv_x}\right)} \frac{VH_x}{\left(1+\frac{r}{100}\right)^{(Freq\_harv_x \times s)-1}} - \sum_{t=0}^{T_x-1}\left(\frac{Mait\_cost_x}{\left(1+\frac{r}{100}\right)^t}\right) \qquad (6)$$

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 $Freq_x \leq T_x$ 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 $TNPV_x$, where $Parcl\_area_x$ is the area (ha^{-1}) of parcel x:

$$TNPV_x = Parcl\_area_x \times NPV_x \qquad (7)$$

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 ($TBiomass_x$). If $Immed\_harv_x = 1$ then,

$$TBiomass_x = Parcl\_area_x \times \frac{Perc\_harv_x}{100} \times Harv\_mass_x \times \text{ru}\left(\frac{T_x}{Freq\_harv_x}\right) \qquad (8)$$

Otherwise, if $Immed\_harv_x = 0$ then

$$TBiomass_x = Parcl\_area_x \times \frac{Perc\_harv_x}{100} \times Harv\_mass_x \times \text{ru}\left(\frac{T_x}{Freq\_harv_x}\right) \qquad (9)$$

and

$$TVolume_x = TBiomass_x \times \frac{1}{BCEF_x} \qquad (10).$$

<u>Example</u>: 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 ¼ 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

Running the Model

Before running the Timber Model, first make sure that the InVEST toolbox has been added to your ArcMap document, as described in the **Getting Started** chapter of this manual. Second, make sure that you have prepared the required input data files according to the specifications in *Data Needs*. Specifically, you will need (1) a shapefile or raster file showing the locations of different timber management zones in the landscape; (2) a table with data on harvest frequency and amount, and the price of timber and cost of harvest; and (3) the discount rate for timber, if other than the 7% US government estimate.

- Identify workspace

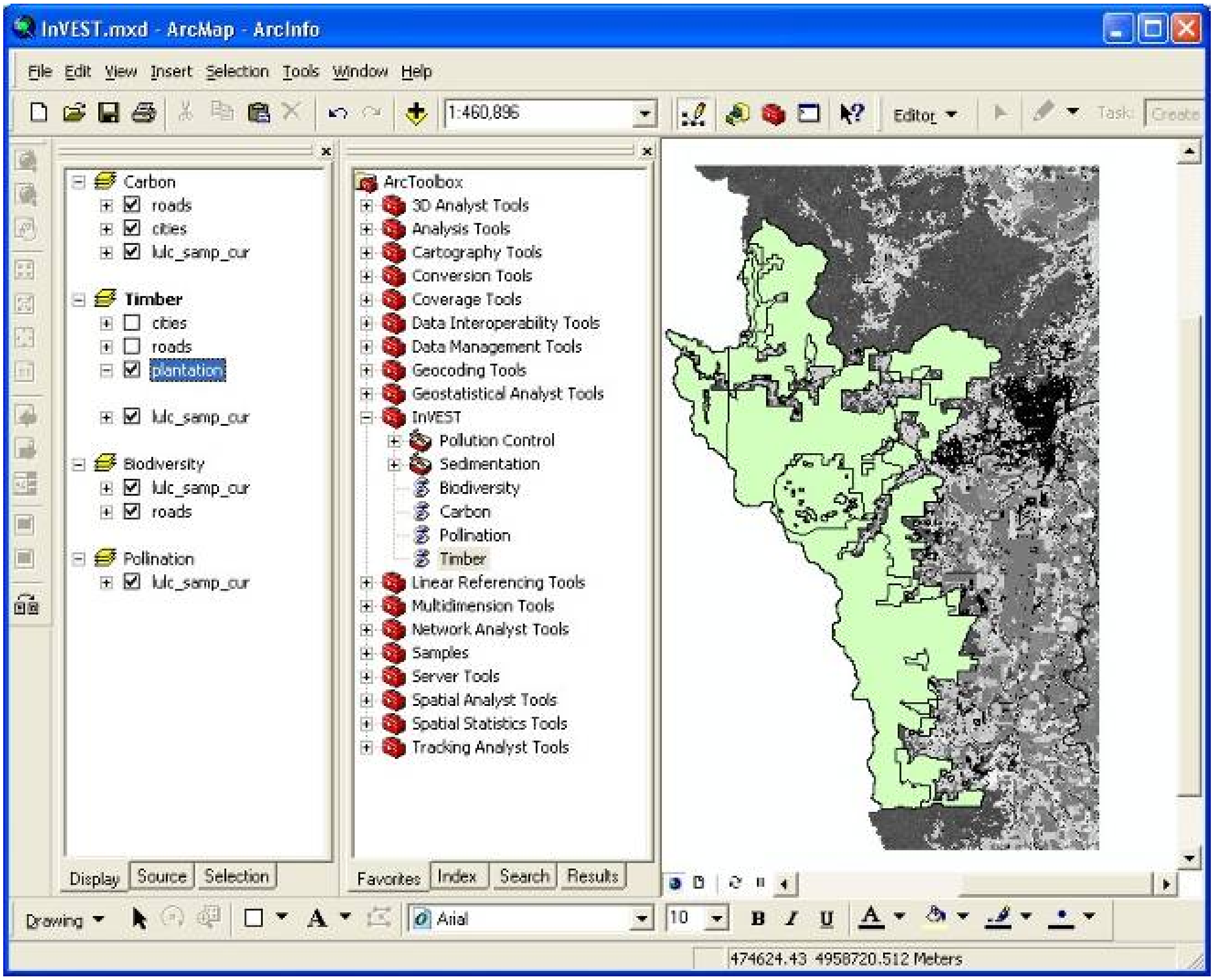
  If you are using your own data, you need to first create a workspace, or folder for the analysis data, on your computer hard-drive. The entire pathname to the workspace should not have any spaces. All your output files will be dumped here. For simplicity, you may wish to call the folder for your workspace "timber" and create a folder in your workspace called "input" and place all your input files here. It's not necessary to place input files in the workspace, but advisable so you can easily see the data you use to run your model.

  Or, if this is your first time using the tool and you wish to use sample data, you can use the data provided in InVEST-Setup.exe. If you unzipped the InVEST files to your C-drive (as described in the **Getting Started** chapter), you should see a folder called /Invest/**timber**. This folder will be your workspace. The input files are in a folder called /Invest/timber/**input** and in /Invest/**base_data**.

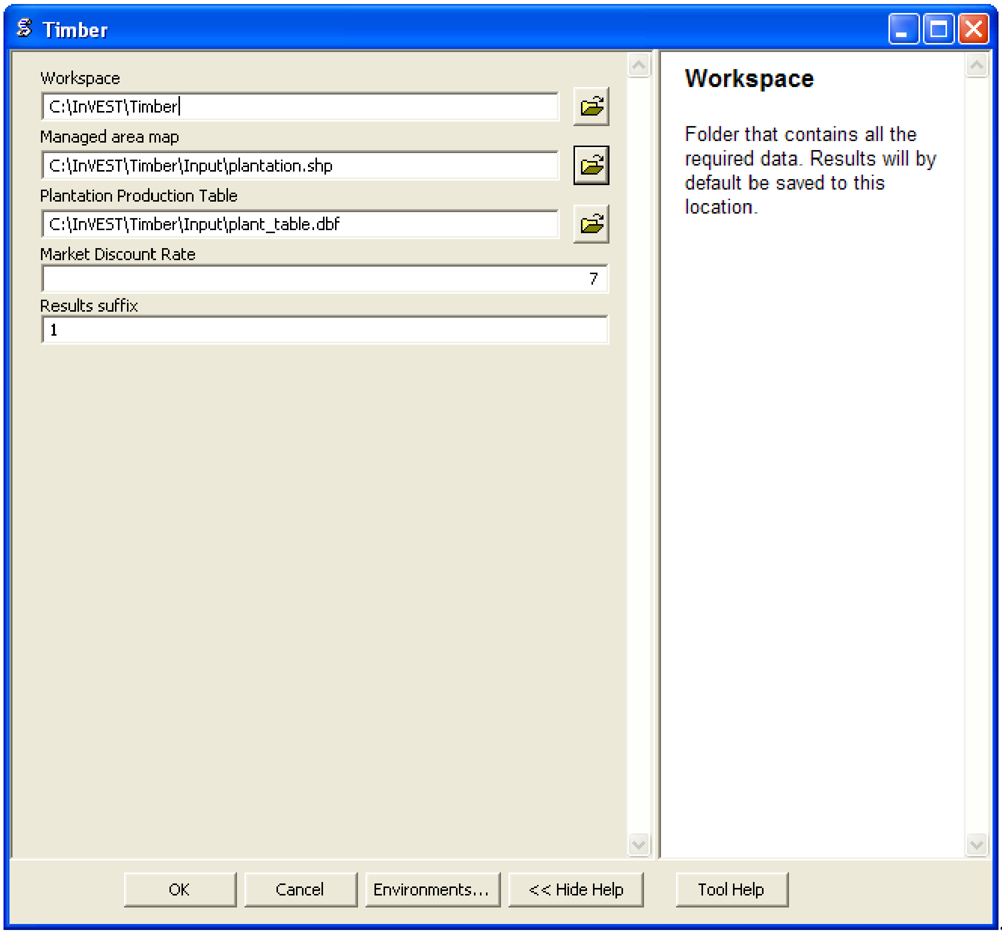
- Open an ArcMap document to run your model.
- Find the InVEST toolbox in ArcToolbox. ArcToolbox is normally open in ArcMap, but if it is not, click on the ArcToolbox symbol. See the **Getting Started** chapter if you don't see the InVEST toolbox and need instructions on how to add it.

- You can run this analysis without adding data to your map view, but usually it is recommended to view your data first and get to know them. Add the data for this analysis to your map using the Add Data button and look at each file to make sure it is formatted correctly. Save your ArcMap file as needed.

- Click once on the plus sign on the left side of the InVEST toolbox to see the list of tools expand. Double-click on Timber.



- An interface will pop up like the one below. The tool indicates default file names, but you can use the file buttons to browse instead to your own data. When you place your cursor in each space, you can read a description of the data requirements in the right side of the interface. In addition, refer to the ***Data Needs*** section above for information on data formats.



- Fill in data file names and values for all required prompts. Unless the space is indicated as optional, it requires you to enter some data.

- After you've entered all values as required, click on OK. The script will run, and its progress will be indicated by a "Progress dialogue."

- Upon successful completion of the model, you will see new folders in your workspace called "intermediate" and "output." These folders contain several raster grids. These grids are described in the next section.

- Load the output grids into 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 file appears in the map.

- You can also view the attribute data of output files by right clicking on a layer and selecting OPEN ATTRIBUTE TABLE.

Interpreting results

Parameter Log

Each time the model is run, a text 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.

Final Results

Final results are found in the output folder of the workspace for this model. The model produces two main output files:

1. Timber_suffix.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 (m^3) 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)).

2. Timber_dateandtime_suffix.txt is a text file that summarizes the parameter data you chose when running the Managed Timber Production 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 can choose in the model's interface (you can also choose to leave it blank).

References

Maass, J., P. Balvanera, A. Castillo, GC Daily, HA Mooney, P. Ehrlich, M. Quesada, A. Miranda, VJ Jaramillo, F. García-Oliva, A. Martínez-Yrizar, H. Cotler, J. López Blanco, A. Pérez Jiménez, A. Búrquez, C. Tinoco, G. Ceballos, L. Barraza, R. Ayala, and J. Sarukhán. 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.

Núñez, D., L. Nahuelhual, and C. Oyarzún. 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.

CROP POLLINATION

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.

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.

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 index on cell *x*, P_x, is simply the product of foraging and nesting such that:

$$P_x = N_j \frac{\sum_{m=1}^{M} F_j e^{\frac{-D_{mx}}{\alpha}}}{\sum_{m=1}^{M} e^{\frac{-D_{mx}}{\alpha}}},$$

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, D_{mx} is the Euclidean distance between cells m and x and α is the expected foraging distance for the pollinator (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:

$$P_{ox} = \frac{P_x e^{\frac{-D_{ox}}{\alpha}}}{\sum_{x=1}^{M} e^{\frac{-D_{ox}}{\alpha}}},$$

where P_x is the supply of pollinators on cell x, D_{ox} is distance between source cell x and agricultural cell o, and α is species' 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, P_o, 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, as:

$$Y_o = 1 - \nu_c + \nu_c \frac{P_o}{P_o + \kappa_c},$$

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:

$$PS_m = \nu_c \sum_{o=1}^{O} V_o \frac{P_{ox}}{P_o},$$

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).

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.

   a. 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).

   b. 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).

c. 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.

Sample data set: \Invest\base_data**lulc_samp_cur**

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):

   a. *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).

   b. 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.

   c. 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.

   d. *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).

<u>Sample data set</u>: \Invest\pollination\input**Guild.dbf**

<u>Example</u>: 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.

   <u>Name</u>: file can be named anything

   <u>File type</u>: *.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.

   <u>Rows</u>: each row is a different LULC class.

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

   a. *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.

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

   c. 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).

   d. 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).

Sample data set: \Invest\pollination\input**LU.dbf**

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

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).
   Sample data set: \Invest\Base_data**lulc_samp_fut**

6. **Proportion of total crop yield attributed only to wild pollination**. This parameter is entered on the interface vary in their dependence on pollinators; some crop species are selfcompatible and yield is less dependent on pollination while other species obligately require pollination to generate any yield (Klein et al. 2007). This model accounts for both observations by using this value. This value would be equal to 1 if a crop is an obligately outcrossing species and equal to 0 if the crop species were wind-pollinated. This parameter is useful if considering a single crop species because it's applied uniformly to all agricultural area.

Running the Model

Before running the Pollination model, make sure that the InVEST toolbox has been added to your ArcMap document, as described in the **Getting Started** chapter of this guide. If you are running

ArcGIS 9.3 then you will also need the GDAL python library to run the pollination model. If you are running ArcGIS 10 then you do not need to add the GDAL library. The installation files are available from the location where you download the InVEST installer but can also be obtained online from the links provided. Below are the installation instructions. These instructions are for Windows XP and may differ for other versions of Windows or other operating systems:

1. Download and install GDAL from the InVEST distribution location (http://invest.ecoinformatics.org) or http://download.osgeo.org/gdal/win32/1.6/gdalwin32exe160.zip.
2. Unzip the GDAL archive into a permanent location (e.g., C:\gdalwin32-1.6).
3. Add your new GDAL bin directory (C:\gdalwin32-1.6\bin, if you installed as above) to your system Path environment variable. To do this, right click on 'My Computer,' 'Properties,' Advanced > Environment Variables. Under system variables, select 'Path' system variable, edit, add a semicolon to separate the existing values then add your GDAL bin directory. For example if the existing Path variable was "C:\Program Files\soft," after editing it should read "C:\Program Files\soft; C:\gdalwin32-1.6\bin" *Do not delete any paths that were there before.*
4. In the same Environment Variables dialog, create a new User Variable named GDAL_DATA with a value of C:\gdalwin32-1.6\data (change this to suit your GDAL install location).
5. Install the GDAL python bindings. Download the package from http://invest.ecoinformatics.org or this location: http://pypi.python.org/pypi/GDAL/1.6.1. Browse to the bottom of that page and select a version that matches your python version.

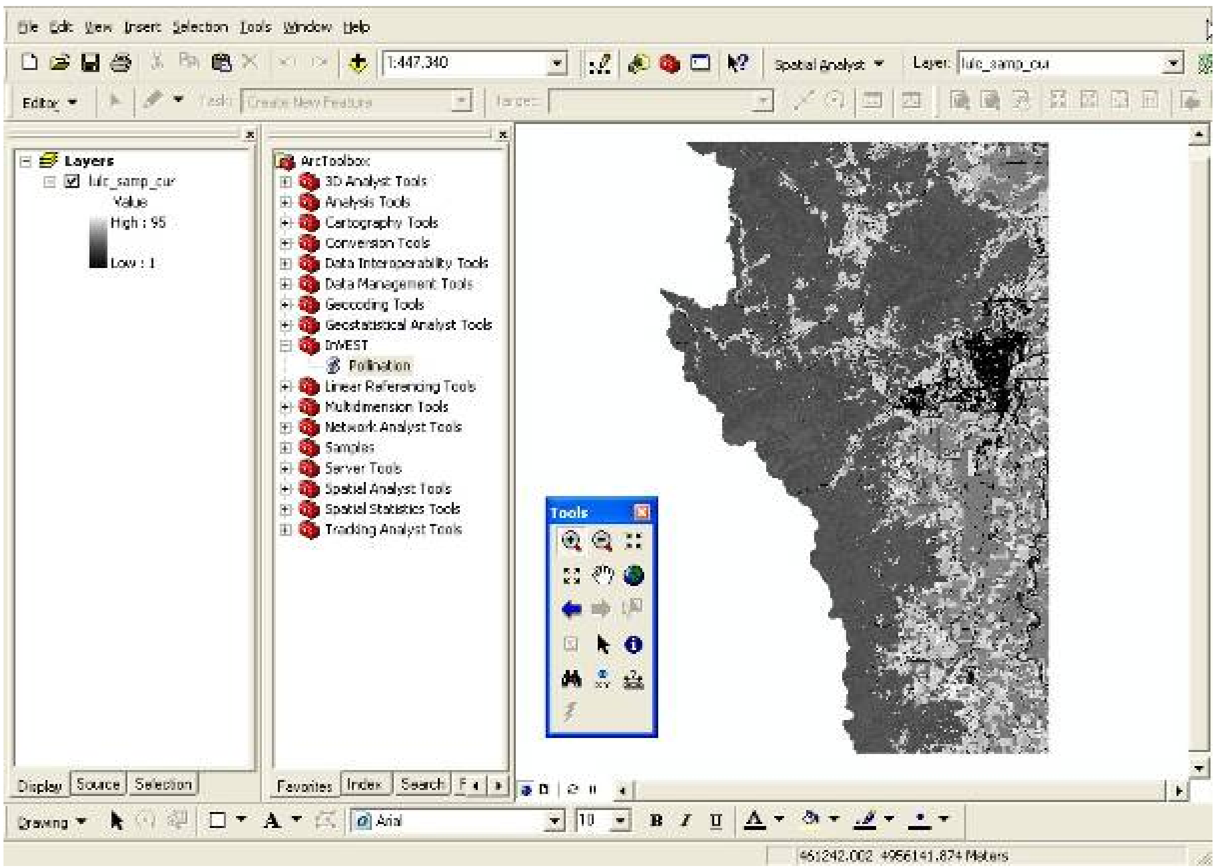
- Make sure that you have prepared the required input data files according to the specifications in *Data Needs*. Specifically, you will need a land cover raster file depicting the different land cover and land use types in the landscape, a Table of Land Cover Attributes, describing the suitability of the land cover types to nesting and floral resources, and a Table of Pollinator Species or Guilds, describing the nesting and seasonal behavior and crop visitation of different pollinators.

- Create a workspace on your computer hard-drive if you are using your data. The pathname to the workspace should not have spaces. All your output files will be dumped here. For simplicity, you could create a folder in your workspace called "input" and place all your input files here. It is not necessary to place input files in the workspace, but this will make it easier to view the data you use to run your model. If this is your first time using InVEST and you wish to use sample data, you can use the data provided in InVEST-Setup.exe. If you unzipped the InVEST files to your C-drive (as described in the **Getting Started** chapter), you should see a folder called /Invest/**pollination**. This folder should be your workspace. The input files are in a folder called /Invest/pollination/input and in /invest/base_data.

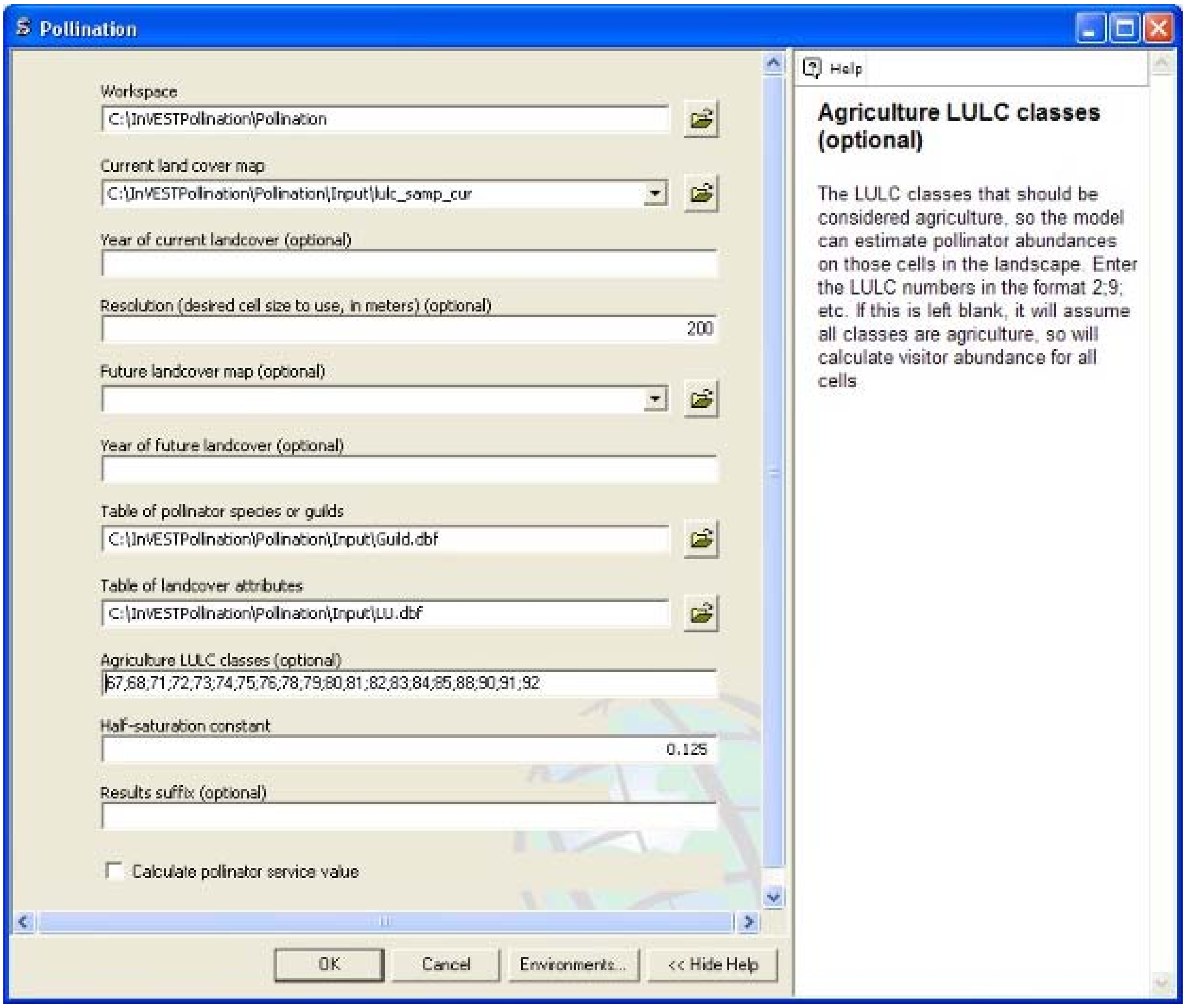
- Open an ArcMap document to run your model.

- Locate the InVEST toolbox in ArcToolbox. ArcToolbox should be open in ArcMap, but if it is not, click on the ArcToolbox symbol. See the **Getting Started** chapter if you do not see the InVEST toolbox.

- Click once on the plus sign on the left side of the InVEST toolbox to see the list of tools expand. Double-click on Pollination.



- An interface will appear like the one below that indicates default file names but you can use the file buttons to browse to your data. When you place your cursor in each space, you can read a description of the data requirements in the right side of the interface. Refer to the ***Data Needs*** section for information on data formats.



- Fill in data file names and values for all required prompts. Unless the space is indicated as optional, inputs are required.

- After entering all required data, click OK. The script will run, and its progress will be indicated by a "Progress dialogue."

- The successful running of the model and the time it takes depends on a combination of the following factors:

    - Size of landscape: If your landscape is very large (e.g., >3 million cells) then you may experience problems. Consider either entering a larger resolution than the original resolution of the image or cropping your image to a smaller extent.
    - Resolution: The cell size chosen for the model run determines the effective number of cells that the model has to handle. Select this carefully depending on the pollinator flight distances.

- o Foraging distances (Alpha): If the Alphas of the pollinators are large (>1000m) then the distance matrix becomes large, which results in a long run time or potential crashing.
- o Number of pollinator species: Since the model processes each pollinator in turn, the more species you have the longer it takes to complete the run.
- o Your computer: The memory and speed of your computer will determine the success and speed of your run. It is preferable to have at least 2GB memory and enough free disk space.
- o On a 3GB memory computer with a 3.5 million cells and 56m resolution, 4 pollinators with alphas between 100m and 2000m the model takes up to 3 hours to run.

- Upon successful completion of the model, you will see two new folders in your workspace called "output" for final maps and "intermediate" for intermediate results. The folders should contain several raster grids, described in the next section.

- Load these grids into ArcMap using the Add Data button. The next section further describes what these files mean. 

- To change the symbology of a layer, right-click on the layer name in the table of contents, select Properties and then Symbology. There are many options to change the file's appearance in the map.

- To view the attribute data of output files, right click a layer and select Open Attribute Table.

Interpreting results

Parameter Log

Each time the model is run, a text file will appear in the output folder. This file lists the parameter values for that run and will be named according to the service, the date and time, and the suffix.

Final results

Final results are found in the "output" folder within the working directory you set up for this module.

Final results are found in the output folder within the working directory set up for this model.

- ***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 *(found in the folder name "intermediate")*

You may also want to examine the intermediate results. These files can help determine the <u>reasons for the patterns</u> in the final results.

- ***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 fil*e name will be the first 4 characters of the 'Species' column in dataset #2, *s*o 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 <u>nesting</u> 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 <u>on each agricultural cell</u> 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.

Appendix: Data sources

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

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/cgi/content/abstract/103/9/1589].

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.

DATA REQUIREMENTS AND OUTPUTS SUMMARY TABLE

Data Requirements for InVEST (2.0)

	Step	Data requirements	Process	Outputs
Habitat Quality and Rarity (Tier 0)				
Required		Land use/land cover	Calculate habitat quality based on threat intensity and sensitivity	Habitat degradation index Habitat quality index
Required		Threat impact distance		
Required		Relative threat impact weights		
Required		Form of threat decay function		
Required		Threat maps		
Required		Habitat preference by species group		
Required		Habitat sensitivity to threats		
Required		Half saturation constant		
Optional		Protected status		
Optional		Baseline land use/land cover	Calculates rarity of each habitat type relative to baseline	Relative habitat rarity index
Carbon Storage and Sequestration (Tier 1)				
Required	Service	Land use/land cover	Looks up carbon stock(s) per pixel	Total carbon stock (Mg/pixel)
Required	Service	Carbon in aboveground biomass		
Required	Service	Carbon in belowground biomass		
Required	Service	Carbon in dead organic matter		
Required	Service	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)
Optional	Service	First year of timber harvest		
Optional	Service	Harvest frequency		
Optional	Service	Half life of harvested wood products		
Optional	Service	Carbon density in harvested wood		
Optional	Service	Biomass conversion expansion factor		
Optional	Service	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	Present value of sequestered carbon (currency/pixel/yr)
Optional	Value	Discount rate		
Optional	Value	Timespan		
Optional	Value	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 evapotranspiration	Annual average water yield (mm/watershed/yr, mm/sub-basin/yr)
Required	Supply	Annual average precipitation		
Required	Supply	Annual average reference evapotranspiration		
Required	Supply	Plant available water content		
Required	Supply	Evapotranspiration coefficient		
Required	Supply	Root depth		
Required	Supply	Effective soil depth		
Required	Supply	Zhang coefficient		
Required	Service	Consumptive use by LULC	Subtracts water consumed for other	Annual average water yield available for hydropower production (mm/watershed/yr, mm/sub-basin/yr)
Required	Service	Watersheds above points of interest		
Optional	Value	Calibration coefficient	Estimates power generated by water available for hydropower	Energy production (KWH/watershed/yr, KWH/sub-basin/yr)
Optional	Value	Turbine efficiency		
Optional	Value	Reservoir fraction for hydropower		
Optional	Value	Average annual head		
Optional	Value	Hydropower production costs	Calculates net present value of energy produced over lifetime of dam	Net present value (currency/watershed/yr, currency/sub-basin/yr)
Optional	Value	Hydropower price		
Optional	Value	Timespan		
Optional	Value	Discount rate		
Water Purification: Nutrient Retention (Tier 1)				
Required	Supply	Land use/land cover	Calculates nutrient export and retention	Nutrient export (kg/watershed/yr, kg/sub-basin/yr) Nutrient retention (kg/watershed/yr, kg/sub-basin/yr)
Required	Supply	DEM		
Required	Supply	Soil depth		
Required	Supply	Water yield (output from Hydropower model)		
Required	Supply	Export coefficient (for nutrient(s) of interest)		
Required	Supply	Nutrient filtration efficiency		
Required	Service	Allowed level of nutrient pollution	Subtracts retention equal to amount of	Nutrient retention of value for water quality (kg/watershed/yr, kg/sub-basin/yr)
Required	Service	Watersheds above points of interest		
Optional	Value	Annual average nutrient removal costs	Calculates present value of costs	Avoided treatment costs (currency/watershed/yr, currency/sub-basin/yr)
Optional	Value	Timespan		
Optional	Value	Discount rate		

Erosion Control (Tier 1)				
Required	Supply	Land use/land cover Rainfall erosivity Soil erodability Crop factor Management factor DEM Sediment retention efficiency Slope threshold	Calculates sediment retention at each pixel using USLE and routing	Annual average erosion (tons/watershed/yr, tons/sub-basin/yr) Annual average sediment retention (tons/watershed/yr, tons/sub-basin/yr)
Required	Reservoir Service	Reservoir dead volume (reservoirs points of interest Watersheds above points of interest	Subtracts sediment equal to dead volume	Annual average sediment retention of value to reservoirs (tons/watershed/yr, tons/sub-basin/yr)
Required	Treatment Plant Service	Allowed sediment load in rivers (TMDL, etc.) Watersheds above points of interest	Subtracts sediment equal to allowed load	Annual average sediment retention of value to water treatment plants (tons/watershed/yr, tons/sub-basin/yr)
Optional	Avoided Dredge Value	Annual average dredge cost Timespan Discount rate	Calculates present value of costs	Avoided dredge costs (currency/watershed/yr, currency/sub-basin/yr)
Optional	Avoided Treatment Value	Annual average sediment removal cost Timespan Discount rate	Calculates present value of costs	Avoided treatment costs (currency/watershed/yr, currency/sub-basin/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 (m^3/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	Calcualtes relative abundance of pollinators	Index of pollinator abundance (realtive abundance/pixel, relative abundance/watershed)
Required	Service	Relative abundance index (supply from above)	Calculates relative 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)

We would like to acknowledge the following sources for data that are provided with the marine 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): http://www.ngdc.noaa.gov/mgg/shorelines/gshhs.html. **(**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 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.

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: Katie Arkema, Gregory Guannel, Anne Guerry, CK Kim, Mike Papenfus, Jodie Toft, Joey Bernhardt, Apollo Yi, Jeremy Davies, Gregg Verutes, Steve Polasky, Mark Plummer, Phil Levin, Bill Labiosa, Francis Chan, Guy Gelfenbaum, Peter Ruggiero, Andre Punt, Melanie McField, Ben Halpern, Sarah Lester, Kai Chan, Malin Pinsky, Mike Beck, Dan Holland, Barry Gold
Others: Matt Marsik, Spencer Wood, Andrew Day, Laura Loucks, Trudy Warner, Kevin Head, Roger Bedard, Jim Regetz