All numbers are fake
This document just
describes a plan for
processing data.

cbm_runner pipeline - Data flowchart
Lucas Sinclair, lucas.sinclair@me.com
Last modified:
Tue Mar 12 2019

1) Input data: Historical inventory data.
- The source is "The National Forest inventories" (NFI).
- A bit different for each country.

Invento...
Invento...
Invento...
Inventory.???

2) Input data: Historical harvest data.
- From 1995 to 2015.
- The source is "Food and Agriculture Organization". Called FAO stat.

harvest.???

Unkown provenance
Manual interventions?

aidt
aidt
aidt
aidb_eu.mdb

5) AIDB-eu: An archive index database specific for each europeen country

Calibration done by R.P.
Manual interventions
?

3) Calibration:
- Contains the back inventory as well as growth curves. 82 tables.

REF_FIS
* Classifer_1_check
* Classifer_2_check
* Classifer_3_check

4) Silvicultural treatment:
- Describes futures disturbances.
- Will be responsible for "dist_events_scenario.csv"

silviculture.sas

calibration.mdb

Assocation
Maps the name of administrative, ecology region, species, disturbances between the calibration.db and the aidb

We only use about 10 tables from this file

Generate new distrubances
Uses information from dist_events_hist and silviculture to create scenario disturbances

Generate CSV files
Converts the calibration database to several CSV files. (Look at upd_aws.py)

Files that have two icons are files that will be used differently for the period before the year 2000 and after.

Coefficients are stored in the calibration.mdb

Versioned

transition_rules.csv
transition_rules_hist.csv

yield.csv
yield_hist.csv

dist_events.csv
dist_events_hist.csv

dist_types.csv

ageclass.csv

classifiers.csv

inventory.csv

associations.csv

coefficients.csv

dist_events_scen.csv   dist_types_scen.csv

Preprocessor
Filter certain row out of the disturbance events

dist_events_filtered.csv

Scenario Genration
Can modify the input files according to current scenarios

Assocation
Create a dictionary of JSON objects. Use a mustache template.

Combine CSV To XSL
Combine the CSVs into a multi-sheet excel file.

() Input XLS: an excel file with seven sheets:

* ageclass
* classifiers
* disturbance_events
* disturbance_types
* inventory
* transition_rules
* yield

input.xls

config.json

() Config JSON: Autogenerated JSON file containing:

* excel file path (c.f. 1)
* output path
* spatial_units
* disturbance_type_mapping
* species_classifier
* nonforest_classifier

StandardImportToolPlugin
- Written in Microsfot Visual Studio C#
- Open source
- Binary distribution

() MS Access database: containing these 121 tables:

tblRootTurnoverKoopit
* tblRunBiomassConversionScenario
* tblRunBiomassTurnoverScenario
* tblRunBiomassTurnoverScenarioLookup
* tblRunClimateScenario

Why are there so many tables given the small input?

project.mdb

() Archive index database: MS Access database containing these 71 tables:

tblGenusTypeDefault
* tblGrowthMultiplierDefault
* tblInput2OutputColMapping
* tblInputDB
* tblInputTable4RPDRBTable

aidb.mbd

CBM-CFS3
- Written in ???
- Closed source
- Dependency on dot net
- Binary distribution

cbm3_python
- Written in Python
- Open source
- Wrapper scripts to call CBM

s.CleanupRunDirectory()
s.CreateMakelistFiles()
s.copyMakelist()
s.runMakelist()
s.loadMakelistSVLS()
s.DumpMakelistSVLs()
s.copyMakelistOutput()
s.CreateCBMFiles()
s.CopyCBMExecutable()
s.RunCBM()
s.CopyTempFiles()
s.LoadCBMResults()

() MS Access database: containing these 51 tables:

tblrystd
* tblLandClasses
* tblManagementType
* tblMaturityType
* tblNIRParcialStock

Why is there a reduction in the number of tables?

project.mdb

Custom Post Processing
Written in Python

coefficients.csv

initialize_carbon_pools

() Updated Inventory: containing the volume of forest of different species and different ages

	Age	Area	Other
1	10	5000	a
2	30	3000	B
3	50	1000	c

() Visulaize: The inventory age according the time step, with area as a third parameter

() Report: Produce a PDF report containing statistics and summarizing the run of the pipeline. This is useful for checking that everything went well.

() Data: Provide access to all the data that was generated

PDF

Orig
Input
Output

Area
0.0
800.0
1600.0
2400.0

AvgAge
TimeStep