BOATS VB1

Guidelines for simulation

Jerome Guiet (jerome.guiet@uab.cat) – 15/04/2017

The BiOeconomic mArine Trophic Size-spectrum (BOATS) model represents the economic activity of fishing explicitly coupled to the ecosystem dynamic. The present document summarizes the structure of the model and provides guidelines for its use. A short tutorial at the end of this document describes the key steps to follow in order to run the model for the first time.

1 – Structure of BOATS VB1 :

BOATS VB1 should be run using Matlab. Any new simulation with the model shoud follow three main steps (see Fig. 1): a pre-processing step; a computation step; a post-processing step. The pre-processing step aims at producing the ecological (*Ecological.mat*) and economical (*Economical.mat*) inputs needed to run the model for distinct scenarios. The computation step solves the governing equations according to the forcing and according to user defined parameters. The post-processing step is a set of scripts allowing the analysis of the output(s) (*Boats_VB1_?.mat*) obtained from simulations.

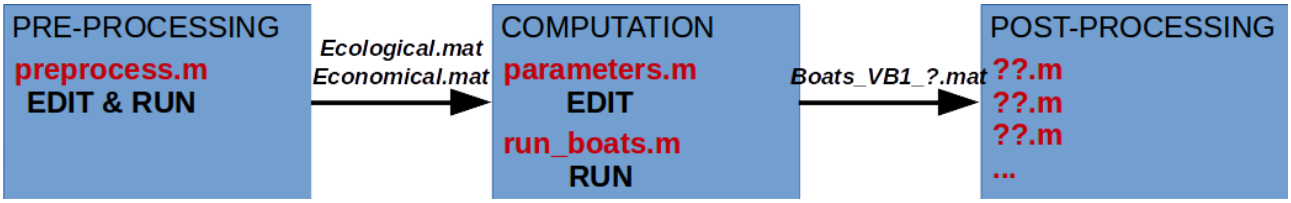

Figure 1: Schematic representation of the structure of the model BOATS VB1 as well as relevant scripts and action.

Each step implies the tuning of paths or parameters in order to adapt the simulation to your scenario or numerical experiment. All modifications for the pre-processing are in the **preprocess.m** script which then needs to be run to generate the model inputs *Ecological.mat* and *Economical.mat*. Then, for computation, all modifications are in the **parameter.m** script, the simulation is completed running the **run_boats.m** script. When a simulation ends successfully an output matlab structure *Boats_VB1_?.mat* is generated and can be analysed using post-processing scripts such as '??.m', '??.m'...

The following sections detail what should be edited in the scripts and grossly explain what is actually happening while the scripts are run.

2 – Pre-processing :

This step defines the domain on which the model will run (0D or 2D). It defines the timeseries of temperature and primary production used to constrain the biology on each computational node in the domain (analytical or empirical/model distributions). It defines the timeseries of fish price, catchability and fishing cost used to simulate different scenarios (analytical or empirical/model distributions).

The data used to force the model come from diverse sources and may not directly satisfy the format accepted by BOATS. In order to run at the computation step the model needs consistency in the inputs. Consistency in the dimensions of the different domains, consistency in the units of the data used as forcing. Ensuring this consistency is the role of this pre-processing step.

Be it in 0D or 2D the BOATS model solves the governing equations on numerical cells, each characterized by diverses ecological and economical conditions. In 0D the model considers only one cell, in 2D it runs on *n* x *m* cells covering part or all the global ocean (see Fig. 2). To consider these domains BOATS needs the following informations:

- A **mask** of the domain: this mask is a grid of the numerical points where the model will run. On this mask, ocean cells should be indicated by 0 values and land cells by 1 values. Figure 2, on the right, illustrates this masking of the 2D simulation domain with a grey shading for cells on land. In 0D the mask indicates a single ocean cell.
- The longitude **lon** and latitude **lat** of the cells in the domain: these coordinates should be two grids, one for the longitude of the center of each cell, one for the latitude of the center of each cell.
- The **surface** of each cell (in m^2) : a grid of the surface of each cell is necessary in order to integrate quantities, such as for example the fish biomass, over the all simulation domain.

Once the simulation domain specified, the model requires at least a biological forcing. In 0D, these are basically time series corresponding with the center of the single cell, in 2D these are sets of maps which will be successivelly used during computation. The necessary forcings are:

- The time serie of net primary production **npp** in each cell (in mmolC m^{-2} s^{-1}): *nt* grids of the net primary production in each cell.
- The time serie of net primary production averaged in the euphotic zone **npp_ed** in each cell (in mmolC m^{-3} d^{-1}): *nt* grids.
- The time serie of **temperature** in each cell (in deg C): *nt* grids.

Finally, when the biology runs coupled with the dynamic of fishing, economical forcings are also necessary :

- The time serie of **price** in each cell (in \$ g^{-1}): *nt* grids.
- The time serie of **cost** in each cell (in \$ W^{-1}): *nt* grids.
- The time serie of **catchability** in each cell (in m^2 W^{-1} s^{-1}): *nt* grids.

The 7 grids/sets of grids **mask**, **lon**, **lat**, **surface**, **npp**, **npp_ed** and **temperature**, are mandatory for any simulation, with or without fishing (see next section). Once properly tuned the **preprocess.m** script generate them and save them in the *Ecological.mat* file. The 3 sets of grids **price**, **cost** and catchability, are only used for simulations with fishing. Once properly tuned the **preprocess.m** script generate them and save them in the *Economical.mat* file.

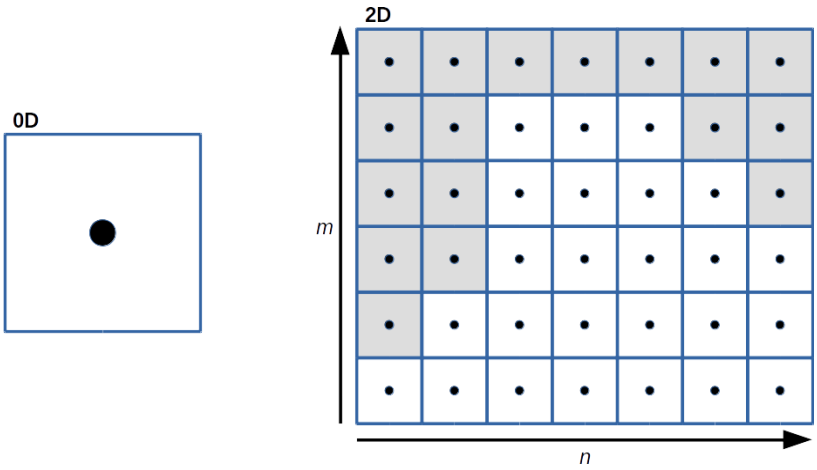

Figure 2: Schematic representation of 0D or 2D (*n x m*) domains. On the 2D domain the grey the shading illustrates the masking of continental cells.

The tunable section of the **preprocess.m** script is identified by the heading 'DEFINE FORCING CHARACTERISTICS'. First, few keys can be edited :

- plot_input : if 1 display few forcing grids for quick control, else no.
- create_ecological : if 1 generate *Ecological.mat* else no.
- create_economical : if 1 generate *Economical.mat* else no.

Then, the minimum informations to generate the simulation domain have to be provided:

- dimfrc : specify the values of the forcing dimensions *m*, *n* and *nt*.
- mask_path : path to find the file where is stored the **mask** grid. This file can be a matlab structure *.mat*, a netcdf file *.nc*...
- mask_var : name of the **mask** variable in mask_path.
- wet / dry : specify how the ocean cells and land cells are identified in mask_var.
- lon/lat_path : path to find the file where is stored the **lon**/**lat** grid. This file can be a matlab structure *.mat*, a netcdf file *.nc*...
- lon/lat_var : name of the lon/lat variable in lon/lat_path.
- surface_path : path to find the file where is stored the **surface** grid. This file can be a matlab structure *.mat*, a netcdf file *.nc*...
- surface_var : name of the **surface** variable in surface_path.
- surface_unit : specify unit of the **surface** in surface_var.

Follow the mandatory information for the biological forcing *Ecological.mat*:

- depth_path : path to find the file where is stored the bathymetry of the simulation domain. This file can be a matlab structure *.mat*, a netcdf file *.nc*... This bathymetry is necessary in order to compute the **npp_ed** forcing.
- depth_var : name of the bathymetry variable in depth_path.
- depth_unit : specify unit of the depth in depth_var .
- ed : specify a selected value for the euphotic zone. This user defined parameter allow the computation of the **npp_ed** forcing.
- npp_path : path to find the file where is stored the **npp** grid. This file can be a matlab structure *.mat*, a netcdf file *.nc*...
- npp_var : name of the **npp** variable in npp_path.
- npp_unit : specify unit of the **npp** in npp_var.
- temperature_path : path to find the file where is stored the **temperature** grid. This file can be a matlab structure *.mat*, a netcdf file *.nc*...

- temperature_var : name of the **temperature** variable in temperature_path.
- temperature_unit : specify unit of the **temperature** in temperature_var.

For coupled simulations with dynamic fishing, you also need to specify the economical forcing *Economical.mat*:

- price_path : path to find the file where is stored the **price** grid. This file can be a matlab structure *.mat*, a netcdf file *.nc*...
- price_var : name of the **price** variable in price_path. The price variable can be user defined (see following), in that case use the 'udef' key.
- price_type : for user defined price forcings if price_var='udef', different basic analytical price distributions can be chosen. See Table 1 for the list.
- price_ref1 : parameter 1 of the user defined price forcing, see Table 1.
- price_dim1 : dimension of the user defined forcing (ongoing improvement...).
- price_unit : specify unit of the **price** in price_var or of the user defined distribution.
- cost_path : path to find the file where is stored the **cost** grid. This file can be a matlab structure *.mat*, a netcdf file *.nc*...
- cost_var : name of the **cost** variable in cost_path. The **cost** variable can be user defined (see following), in that case use the 'udef' key.
- cost_type : for user defined cost forcings if cost_var='udef', different basic analytical cost distributions can be chosen. See Table 1 for the list.
- cost_ref1 : parameter 1 of the user defined cost forcing, see Table 1.
- cost_dim1 : dimension of the user defined forcing (ongoing improvement...).
- cost_unit : specify unit of the **cost** in cost_var or of the user defined distribution.
- catch_path : path to find the file where is stored the **catchability** grid. This file can be a matlab structure *.mat*, a netcdf file *.nc*...
- catch_var : name of the **catchability** variable in catch_path. The **catchability** variable can be user defined (see following), in that case use the 'udef' key.
- catch_type : for user defined catchability forcings if catch_var='udef', different basic analytical catchability distributions can be chosen. See Table 1 for the list.
- catch_ref1 : parameter 1 of the user defined catchability forcing, see Table 1.
- catch_ref2 : parameter 2 of the user defined catchability forcing, see Table 1.
- catch_dim1 : dimension of the user defined forcing (ongoing improvement...).
- catch_unit : specify unit of the **catchability** in catch_var or for the user defined distribution.

Once the parameters of the **preprocess.m** script tuned, run it and follow the instructions eventually displayed. Figures of the domain and mean forcing will be eventually displayed. Once it completes successfully, you should have created either the *Ecological.mat* file, the *Economical.mat* file, or both depending on what you selected. The BOAT model is now ready to compute !

Table 1: User defined types of economical forcings for X=price/cost/catch and necessary parameters. The associated equation is also detailed.

X_type	X_ref1	X_ref2	X_ref3	Equation
'cst'	x	-	-	X_var(t)=X_ref1
'rate'	x	x	-	X_var(t0)=X_ref1 X_var(t)=X_var(t-1)*(1+X_ref2)$^{1/12}$
...				

3 – Computation :

At this step, the governing equations are solved on the domain and with the inputs provided by pre-processing. Many paramaters can be tuned for the computation, for instance in order to modify the number of time steps of the simulation, or in order to modify some of the biological characteristics of modeled fish communities, among others...

An important parameter to edit is the simulation type sim_type. Depending on the numerical experiment you wish to complete, the model can run without or with coupling of the dynamic of the fish biomass with the dynamic of fishing. This is respectively specified by sim_type='nh' (no harvest) or 'h' (harvest). Any BOATS simulation is usually constituted of firstly a no harvest ('nh') run, in order to initialize the biomass distribution in the region of interest, then a harvest ('h') run which is initialized with the biomass distribution estimated by the 'nh' run (see figure 2).

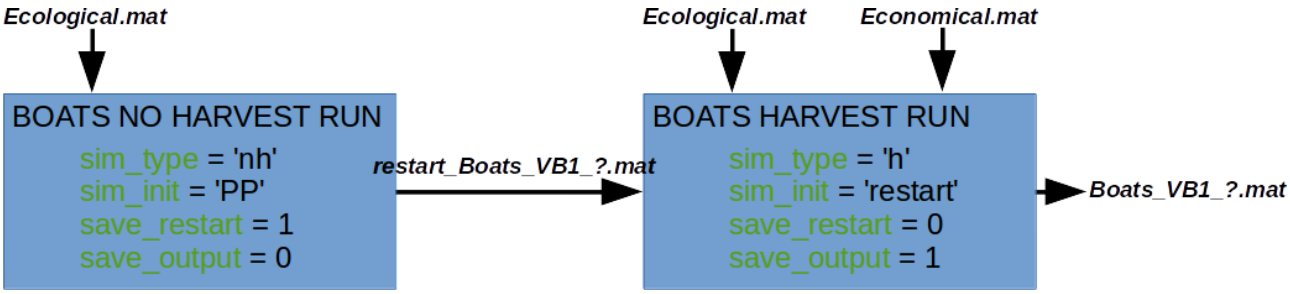

Figure 2: Schematic representation of a typical BOATS simulation, a no harvest ('nh') computation before a harvest ('h') one.

The tunable parameters are all summarized in the **parameters.m** script. The main parameters to modify are identified by the heading 'MAIN TUNABLE PARAMETERS'. These first include paths:

- wrkdir : path to the working directory of the model.
- outdir : path where to direct the *Boats_VB1_?.mat* and restart_*Boats_VB1_?.mat* output files generated at the end of a simulation.

Then parameters which harvest is considered or not and parameters that turn on/off the generation of the *Boats_VB1_?.mat* and *restart_Boats_VB1_?.mat* output files:

- sim_type: no harvest 'nh' or harvest 'h'.
- sim_init: at the beginning of computation, the biomass in the computation domain is initialized with an analytical distribution, option 'PP', or with a distribution coming from a previously generated restart file, option 'restart'.
- save_restart: generate a *restart_Boats_VB1_?.mat* file at the end of the simulation, yes=1, no=0.
- save_output: generate a *Boats_VB1_?.mat* file at the end of the simulation for post-processing analysis, yes=1, no=0.

As illustrated figure 2, usually a simulation follows two steps. First a run without harvest ('nh') initialized with an analytical field ('PP') and forced by the ecological forcings *Ecological.mat*. This run generates the biomass distribution in the domain and save it as restart file *restart_Boats_VB1_?.mat*. Second a run with harvest ('h') initialized with the previously generated restart file ('restart'), forced by the ecologial *Ecological.mat* and the economical *Economical.mat* forcings. This run computes the coupled biomass and fishing dynamic and save it in an output file *Boats_VB1_?.mat*.

The last essential parameters to specify are about the length of the simulation:

- run_length: the number of years for a simulation.
- dtt: the number of days per time step.

Note that the remaining parameters in **parameters.m** are default values for all constants that the model needs. They can be tuned by the user to fulfill particular needs, but for the first runs, it is advisable to leave these parameters and use different combinations of parameters provided in a separate file: *ensemble_parameters.mat*. In *ensemble_parameters.mat*, 5 sets of biological parameters fulfilling precise criterias are provided. So far they represent the best guess to obtain "realistic" simulations. In order to use these sets of parameters, the key param_ens should be set to 1 and the dataset name dataset_ens='ensemble_parameters.mat'.

Once the parameters properly defined for a run without or with harvest, run **run_boats.m** in order to complete your simulation. The model will complete one run of run_length years, or if ensembles are used for forcing, 5 runs of run_length years. The model follows the steps:

- It updates and modifies parameters in the function modify_parameters(): if ensembles are chosen instead of the default set of parameters defined in **parameters.m**, the model updates the parameter values to the corresponding ensemble.
- It loads the forcings in the function load_forcing(): the pre-processed inputs *Ecological.mat* and eventually *Economical.mat* are loaded.
- It prepares the outputs format in the function initialize_output(): a set of prognostic variables are extracted during simulation in order to be saved as output. There characteristics are defined in this function (To improve).
- It sets different variables and structures in the function set_structure(): this function essentially prepares different variables in order to improve the computation performance of the model.
- It initialize the biomass distribution in the domain in the function initialize_domains(): the biomass distribution at the beginning of the simulation is either determined analytically or determined loading a previously generated *restart_Boats_VB1_?.mat* file.
- It computes and solves the equations in the function integrate(): the governing equations are numerically solved and the model runs over run_length years. The detail of the equations is in the papers: D.A. Carozza, D. Bianchi and E.D. Galbraith (2015), The ecological module of BOATS-1.0: a bioenergetically-constrained model of marine upper trophic levels suitable for studies of fisheries ans ocean biogeochemistry, Geoscientific Model Development, 8:10145-10197 ; D.A. Carozza, D. Bianchi and E.D. Galbraith (2017), Formulation, General Features and Global Calibration of a Bioenergetically-Constrained Fishery Model, PLOS ONE, 12:1-28.
- It save a restart file in the function save_restart(): once the integration completed the computation is finished, the model eventually save a restart file if save_restart was set at one.
- It save a output file in the function save_output(): once the computation is finished the model eventually save an output file which will be analyzed at the post-processing step, if save_output is set at one.

The model will follow the same steps in case of no harvest ('nh') or harvest ('h') simulations.

Once a simulation completes sucessfully you should find *Boats_VB1_?.mat* files generated in the output directory outdir. Lets look at them in post-process !

4 – Post-processing :

A set of basic tools and scripts should be available soon in order to browse through the data, but you can already look at the outputs yourself...

5 – First run :

This last section describes the steps to follow when runing BOATS for the first time:

Step 1: Download and unzip the *BOATS_VB1.zip* tarball in the folder you wish to work in:
> unzip BOATS_VB1.zip

Step 2: Open matlab and open the preprocessing script **preprocess.m**.

Step 3: Specify the paths, and units of your forcings in the section 'DEFINE FORCING CHARACTERISTICS'. The tarball comes with forcings of a domain covering the Scotian Shelf in the North Atlantic (in */input*), for testing. Normally you don't need to change anything.

Step 4: Run the **preprocess.m** script. If everything goes smoothly few figures should appear illustrating the domain and forcings around the Scotian Shelf. Once it runs sucessfully, the *Ecological.mat* and *Economical.mat* inputs should be generated in your working directory.

Step 5: Open the **run_boats.m** and **parameters.m** scripts.

Step 6: Specify the working directory workdir on your machine and the output directory outdir in **parameters.m**, in the 'MAIN TUNABLE PARAMETERS' section.
> boats.param.path.wrkdir = ['/my/workdir/directory/'];
> boats.param.path.outdir = ['/my/outdir/directory/'];

Step 7: First, adapt the parameters in the 'MAIN TUNABLE PARAMETERS' section in order to realize a simulation without harvest ('nh'). The objectiv of this first run is to determine the biomass distribution from the environmental forcing *Ecological.mat*:
> boats.param.main.sim_type = 'nh';
> boats.param.main.sim_init = 'PP';
> boats.param.main.save_restart = 1;
> boats.param.main.save_output = 0;
> boats.param.main.run_length = 200;
> boats.param.main.dtt = 30;
You can edit other parameters if you wish...

Step 8: Run the no harvest simulation running **run_boats.m**. It may take a while, eventually at the end you will have 5 restart files *restart_Boats_VB1_?.mat* for the five ensembles of parameters simulated by the model.

Step 9: Second, adapt the parameters in the 'MAIN TUNABLE PARAMETERS' section in order to realize a simulation with harvest ('h'). The output of this simulation will allow the analysis of how fishing and biomass production dynamically interact:
> boats.param.main.sim_type = 'h';
> boats.param.main.sim_init = 'restart';
> boats.param.main.save_restart = 0;
> boats.param.main.save_output = 1;
> boats.param.main.run_length = 200;
> boats.param.main.dtt = 30;
You can edit other parameters if you wish...

Step 10: Run the harvest simulation running **run_boats.m**. It may take a while again, eventually at the end you will have 5 output files *Boats_VB1_?.mat* for the five ensembles of parameters

simulated by the model.

Step 11: That is it, you completed you first simuation of the coupled dynamic of fish biomass and fishing with BOATS, on the Scotian Shelf. You can open the ouput fils in matlab and browse through the data, for example :

```
> load boats2d_VB1_h_ind_6363.mat
```

HAVE FUN !!!