

# *Calibration of ABM on multiple datasets and simulation set-up uncertainty*

*Clément Viguié*

*November 13, 2017*

Calibration of process-based models is a crucial step in order to gain confidence in the model and test the stability of model outputs. Ecological models tend to include a lot of processes and consequently a large number of parameters and outputs. It is necessary in this case to have multi-dimensional observation data to compare the model's outputs to, within the calibration procedure. However, ecological data are rarely exhaustive on all dimensions (question focused, destructive analysis and large cost). Using multiple sources of data may compensate for this lack of detailed description of the system, but come with their own problem.

Calibration of process-based models is a crucial step in order to gain confidence in the model and test the stability of model outputs. Ecological models tend to include a lot of processes and consequently a large number of parameters and outputs. It is necessary in this case to have multi-dimensional observation data to compare the model's outputs to, within the calibration procedure. However, ecological data are rarely exhaustive on all dimensions (question focused, destructive analysis and large cost). Using multiple sources of data may compensate for this lack of detailed description of the system, but come with their own problems. Two main difficulties can be identified here:

- the inability to use the model's parameters to capture part of the uncertainty on the simulation set-up since there are multiple set-ups (one for each source of variation);
- ?weight the different sources/outputs tested.

To these obstacles, in the particular use of this generalizing model, the uncertainty around the species parameters adds up.

## *How to calibrate such model ?*

### *Non divided uncertainty*

One approach of the calibration can be to consider all parameters as equal, and calibrate for each observation data all parameters at once, then select all overlapping sub-space in the parameters hypercube. However because we have different datasets, uncertain simulation and species specific parameters will not always be the same. How do we cope with that ?

### *Multi-level uncertainty*

The objective of calibration is to reduce uncertainty about model parameters. To do that models outputs are compared to observation data. In case of insufficient description of the experiment design, one can make an assumption on the design and use arbitrary values for the uncertain simulation parameters. This can decrease the goodness of fit of the test model's parameters since the conditions are con-

sidered certain and greatly reduce the chances of convergence. The other option is to consider these simulation parameters to be uncertain and simulate with different values. Doing so information contained in the observation data may be lost in the simulation set-up uncertainty. What needs to be done is the reduction of the simulation settings parameters and specific parameters uncertainty in order to the maximum information to be captured in the model parameter selection.

For each of the observation data the maximum information should be gathered and inject in the set-up step. Remaining parameters (uncertain) draw random values into the possible ranges.

*Filtering* Intuitively, regarding only one observation value, to be correct a parameter vector should be able to reproduce it at least once within the simulations drawn for the small uncertain simulation-species parameter space. If conditions for the simulation are highly uncertain, this will not be selective enough. On the other hand, the tighter the distribution of simulations outputs around the observed value denote a little influence of the settings of the simulation over the output and do not justify their existence as settings (as opposed as parameters that are considered fixed in the context of the model).

What is a good criterion for filtering ? See Hartig. Rejection filter can be based on the variability created by the model or estimated uncertainty of the observed data. In my case, the variability of the model outputs will emerge from model little stochasticity and uncertainty of settings that can be related to uncertainty of observed data (we are sure about the data, but not certain how it is linked to experimental conditions that are unknown). This criterion can be set and changed latter (in that case need to store the distribution of outputs for each parameter sample).

*Combining* As described the calibration algorithm will compute acceptance of each model's parameter values sample for each observation/experiment simulation results are compared with. To give value to the parameter value and define an acceptable parameter subspace, one can select the sample that maximize the number of acceptance. To discriminate between samples from a same subset (samples with the same number of acceptance), one could also give weight to any of the observation based on the importance of the pattern or the confidence is the observation.