Agent-Based Modeling and Simulation

Patterns for Parameterization and Calibration

Dr. Alejandro Guerra-Hernández

Instituto de Investigaciones en Inteligencia Artificial
Universidad Veracruzana
Campus Sur, Calle Paseo Lote II, Sección Segunda No 112,
Nuevo Xalapa, Xalapa, Ver., México 91097

mailto:aguerra@uv.mx
https://www.uv.mx/personal/aguerra/abms

Doctorado en Inteligencia Artificial 2024

Credits

- These slides are based on the book of Railsback and Grimm [1], chapter 20.
- Any difference with this source is my responsibility.
-
-

Universidad Veracruzana

Parameterization

- Parameters are the constants in the equations and algorithms we use to represent the processes in an ABM.
- Example. In the butterfly's virtual corridors model, the parameter q represented the probability that a butterfly would deliberately move uphill at a given time step.
- Parameterization is the selection of values for a model's parameters.
- Few of the models we've used are realistic representations of real systems, so we couldn't say what a good value for them is.
- But now, POM is just about relating our models to real systems.

Universidad Veracruzana

Justification

- Parameterization is important because quantitative results matter.
- The ultimate task of ABM is to understand the organization and dynamics of complex systems, including the relative importance of the different processes that we include in our ABM.
- Changing parameter values can change the relative importance of different processes and, hence, the organization and dynamics of the model.
- We must use empirical information from real systems to anchor parameter values to reality, so we can:
  - Learn which processes are really most important and
  - Which dynamics are most believable.

Universidad Veracruzana

Calibration

- Calibration adjusts few specially important parameters by seeing what values cause the model to reproduce patterns observed in the real system, *i.e.*, POM.
- Now, however, these patterns are typically quantitative instead of qualitative and more descriptive of the whole system, not its agents and parts.
- Calibration comes after we've identified theory for behavior and assembled the full model.
- Now we are focused on quantitative results of the full model and comparing them to data from a specific real system.

Universidad Veracruzana

Purposes

- To force the model to match empirical observations as closely as possible, assuming that the adjusted parameters then produce more accurate results when modeling any conditions.
- To estimate the value of parameters that we cannot evaluate directly. We can estimate a value inversely by adjusting it until the model best matches some observations.
- To test a model's structural realism: Can we calibrate it to match the observations within a reasonable range? Or is there something wrong with the model so that it cannot be forced to match observations closely?

Universidad Veracruzana

Learning Objectives

- Understand the several objectives of calibration, and how parameterization and calibration of ABMs differs from calibration of more traditional models.
- Understand some fundamental calibration concepts and strategies, such as identifying good calibration variables and selecting specific criteria to measure a model's calibration.
- Develop experience by conducting several calibration exercises.

Universidad Veracruzana

Traditional modeling

- For many decades, in many fields, modeling to many people has been a process of selecting one or a few simple equations with a few parameters, and then setting the value of those parameters by calibrating the model to empirical data.
- Parameterization then is equivalent to calibration, and more parameters often means more uncertainty because the same limited amount of empirical data must be used to estimate them.
- Most of the information about the system is in the calibrated parameter values because the model contains very few mechanisms or processes.
- Each parameter is an extremely simple sub-model because it represents a process that can actually be quite complex

Universidad Veracruzana

The ABM approach

- ABMs, in contrast, contain more information about the system because they use more entities, state variables, and sub-models.
- And each sub-model can be parameterized and tested by itself.
- Therefore, even though ABMs typically have more equations and parameters than simple models, they are typically less reliant on calibration.
- Calibration of an ABM is often only a matter of fine-tuning a small fraction of its parameters.
- Indeed, many fairly complex ABMs have proven useful and interesting with little or no calibration of the full model.

Universidad Veracruzana

Strategies

- One of the major strategies for making ABMs reliable and credible is to develop, parameterize, and test each sub-model independently and thoroughly.
- If your ABM is said to have too many parameters to calibrate reliably, you need to show how you evaluated most them before even thinking about calibrating the whole model.
- This is about strategies for choosing values for as many parameters as possible, especially in sub-models, without falling back on calibration to evaluate them inversely.

Universidad Veracruzana

Use Published Models

- This task is often similar to parameterization of traditional models.
- One solution is to select, as your sub-model for some process, a published model (already described, calibrated, and analyzed).
- You must of course be sure that it is appropriate for your ABM, especially its time and space scales.

Universidad Veracruzana

Make up Your Own Sub-models

- Otherwise, appropriate parameter values may be available from the literature, your own data, or even via educated guesswork.
- One important trick for credibility is to design your model so its parameters represent real quantities or rates that could, at least in principle, be measured empirically.
- Otherwise, your only choices are guesswork and calibration, and it will be very difficult to show that your values are realistic.

Universidad Veracruzana

Causality

- Another trick is to show that your parameter values cause the sub-model to produce reasonable results in all situations that possibly could occur.
- Do this as part of sub-model development.

Universidad Veracruzana

Use the Traditional Methods, but...

- Sometimes it is necessary to calibrate a sub-model to data, using methods more similar to classical model fitting and calibration.
- It could be wise to consult with modelers familiar with the model fitting literature.
- But, for the first, simple versions of your models, or for more theoretical models designed only to explore possible explanations, it is sufficient to use guesstimated parameter values.

Universidad Veracruzana

Guestimating

- Define upper and lower bounds for the parameter's value, beyond which the sub-model would produce nonsense.
- Then, think about whether the process being modeled operates on a fast or slow time scale, has a strong or moderate impact, occurs very often or rarely, varies a lot or not, etc.
- This helps defining a reasonable range for the parameter, from which you can simply select the middle or try several values.
- Analyze the sub-model and make sure that it works with your guesstimated values before putting it in the model.
- You should force yourself to set the value of every sub-model parameter here, as best you can.

Universidad Veracruzana

Ideas

- Calibration of a full ABM is used to improve and documents its accuracy and to evaluate its most uncertain parameters.
- The basic idea of calibration is to execute a model many times, each time using different values of its parameters, and then analyze the results to see which parameter values caused the model to best reproduce some patterns observed in the real system.
- In any field of science where modeling is common, there is an extensive and important literature on model calibration and model fitting, addressing such topics as mathematical measures for how well model results fit an observed data set.

Universidad Veracruzana

Observations I

- There is only a limited amount of information in the observed patterns and data we calibrate a model to, so using the same information to calibrate more parameters, makes our estimates less certain.
- By fine-tuning many parameters, we can cause a model to match a few patterns very closely while actually making the model less accurate in general, *i.e.*, over fitting.
- The data sets we use to calibrate models always have their own errors and uncertainties, and always represent one unique set of conditions.
- If we try too hard to make the model reproduce one data set, the chosen parameter values would be not as good at representing all the other conditions.

Universidad Veracruzana

Observations II

- There is a cost to calibrating more parameters: If we decide to adjust three instead of two parameters, we are likely to fit the calibration data better, but we need to realize that we are also reducing the certainty in our parameter estimates and increasing the potential for overfitting.
- We need to use, if possible, techniques such as validation via secondary predictions and robustness analysis to assess how well the calibrated model represents the system in general.
- Here are some techniques for doing calibration effectively, presented as a sequence of six steps.

Universidad Veracruzana

Criteria

- Identify a very small number of parameters to evaluate via calibration.
- These should be both especially uncertain and important, and each should have relatively independent effects on the model.

Uncertain Parameters

- A parameter is uncertain if we do not know its correct value accurately, only that it might fall within some broad range;
- Or also because we use it to represent something that is actually variable and complex, and we do not want that variability and complexity in the model.
- Hence, there is no single correct value of the parameter, and it might even be unmeasurable.

Important Parameters

- How do we know which parameters are especially important to a model? This question is only answered convincingly by sensitivity analysis.
- It is often prudent to reconsider model calibration after conducting a sensitivity analysis.
- But even before the sensitivity analysis, we usually have a good idea about which parameters strongly affect model results.
- Run simple preliminary sensitivity experiments, varying the especially uncertain parameters to see how they affect results.
- A parameter that is highly uncertain but has little effect on results should not be used for calibration.

Universidad Veracruzana

Redundancy

- You should avoid having two or more calibration parameters that have similar effects.
- Example. Imagine a business model with two uncertain parameters:
  - The probability of failure due to bankruptcy and
  - The probability of a business closing voluntarily.

  When calibrating, it could be very hard to distinguish the effects of these two parameters: many combinations of them could make the model reproduce the data equally well.
- You could, however, calibrate one of the parameters while assuming the other is a constant percentage of the first.
- Example. Voluntary closure is always 30% of bankruptcy.

Universidad Veracruzana

Categorical Calibration

- You search for parameter values that produce model results within a category or range you have defined as acceptably close to the data.
- Example. You could determine from the calibration criteria that the model is adequately calibrated if the mean number of agents is between 120 and 150 and if their mean size is between 210 and 240.
- You could then run the model with many combinations of parameter values and see which parameter combinations produce results within the acceptable category.

Universidad Veracruzana

Best-Fit Calibration

- You search for one set of parameter values that cause the model to best match some exact criteria, essentially, an optimization.
- Example. You would try to find the parameter values that cause model results to be as close as possible to, for example, mean number of agents $= 135$ and mean agent size $= 225$.

Universidad Veracruzana

Choice I

- Which approach is best?
- Best-fit calibration has the advantage of producing a simple, single set of calibrated parameter values instead of a whole set of acceptable values.
- You can always find a "best" set of parameter values, even if the fit they produce between model and data is not very impressive.
- However, to do this you must be able to identify just one measure of model fit that the best parameter values optimize: you cannot optimize several criteria at once.

Universidad Veracruzana

Choice II

- Categorical calibration requires less interpretation of model results because you are not trying to find absolutely best parameter values, and makes more sense if the calibration data are themselves uncertain.
- But it is always possible that categorical calibration does not solve the problem of telling you what parameter values to use for future model runs because no parameter combinations produce results within the acceptable ranges.

Universidad Veracruzana

Ideas

- Seems natural, since ABMs simulate how a system changes over time, and we often have data collected over time.
- If our model's purpose includes representing how results change over time, then it makes sense.
- Example. How long does it take the system to recover from some perturbation?
- It is not useful when the ABM is intended to explain long-term average conditions, so they intentionally do not contain all the processes that cause the real system to change over time and use no input data to represent how the agents' environment changes over time.
- Example. The Woodhoopoe model.

Universidad Veracruzana

No Time-Series Calibration

- When we have a time series of observations but choose not to calibrate the model's ability to reproduce changes over time, we can instead calibrate the model by its ability to reproduce statistical characteristics of the time series.
- Example. The mean number of agents.
- We often also want to calibrate some measure of variability over time.
- Example. The standard deviation in the time-series data: Does our model produce the same degree of variation around the mean as we see in the observed patterns?

Universidad Veracruzana

Time-Series Calibration I

- When we do choose to use time-series calibration, the question is how to quantify the fit between model results and observations over time.
- We want the mean and standard deviation in model results over time to be close to those in the observed time series...
- But these measures do not reflect trends over time. Some other relatively simple measures for this include:

  Maximum error. The maximum absolute difference between model results and observations, over all the times in the series. The best calibration would produce the smallest maximum error.

Universidad Veracruzana

Time-Series Calibration II

Mean squared error. The mean, over all times in the series, of the square of the difference between model results and observations. Squaring the differences makes sure that negative errors do not offset positive errors, and emphasizes large errors more than small ones.

Accuracy. The number of points in the time series where model results are acceptably close to observations. The modeler must decide what range around the observed values is close enough to consider a successful calibration.

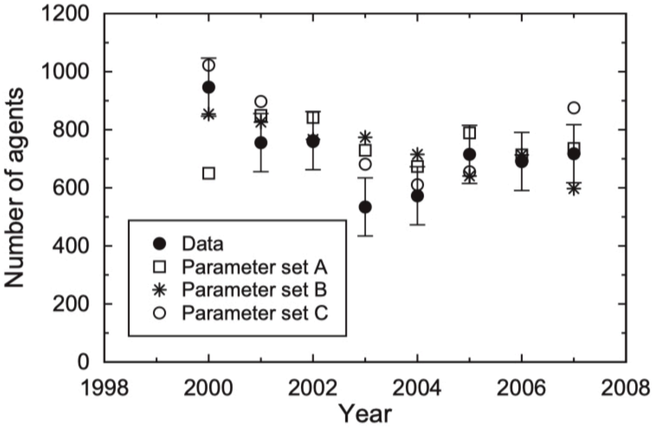
Universidad Veracruzana

Example

- Assume a model that is being calibrated with data on the number of agents alive or active at the end of a year, observed for eight years from 2000 to 2007.
- We can plot these data plus results from three model runs using three sets of parameter values.
- Assume that the model can be considered adequately accurate (well calibrated) if its results are within 100 agents of the observed data each year.
- In the following graph, can you tell from the figure which parameter set is best?

Universidad Veracruzana

The plot


1200
1000
800
600
400
200
0
Number of agents
1998
2000
2002
2004
2006
2008
Year
Data
Parameter set A
Parameter set B
Parameter set C

Observations

- In some years, *e.g.*, 2001, 2003, none of the parameter sets produced results close to the data.
- For the first few years parameter set B seems to work well, but in the last few years set A seems to be best.
- The measures of model fit discussed above are evaluated for these results:

Measure	A	B	C
Difference in mean	36	**24**	62
Difference in sd	53	38	**17**
Maximum error	296	240	**158**
Mean squared error	19,782	13,976	**9,699**
No Years within 100	5	5	5

Universidad Veracruzana

Calibration Criteria

- This step is to identify and define the quantitative patterns that will be used to calibrate the model.
- This actually is often the most challenging part of calibration.
- We have to identify some key patterns that we calibrate the model to reproduce, and define them, and the model results that we will compare them to, with sufficient quantitative precision so we can say mathematically whether the pattern is met or how close the model results are to the pattern.

Universidad Veracruzana

Selection

- First, we need to calibrate an ABM against patterns in the different kinds of results that will be used in the model's final application.
- Example. Is the model going to be used to explain the number, size, and location of agents? If so, then it is best to calibrate it against patterns in all these kinds of results if we can.
- The purpose strikes back!

Universidad Veracruzana

Comparability

- Second, we need to make sure that observed patterns and model results used for calibration actually represent the same system characteristics:
  - The observations need to be from a system that has the same basic mechanisms that the model does, and collected under conditions similar to the ones represented in the model.
  - Observations and model results need to measure the same things at the same times, as exactly as possible.
  - The observations should have been made at spatial and temporal resolutions compatible with the model's time step and grid size.
  - Special care must be taken using measures of variability as calibration criteria. There are many kinds and causes of variability, and it is not legitimate to compare one kind of variability to a different kind.

Universidad Veracruzana

Accuracy of the Selected Patterns

- Third, How reliable and accurate the patterns are?
- All observations have errors and uncertainties, and we often must calibrate a model using data that we know could be inaccurate.
- Using uncertain data is usually unavoidable: we don't worry so much about matching them exactly when facing more uncertainty, and we recognize that calibrated parameter values are less certain.
- But we need to have at least some idea of how accurate or certain the observations are, so we know how much information they really contain.

Universidad Veracruzana

Comparison

- Finally, we must specify how we will compare the observed patterns to model results, to determine which model runs produced results that are calibrated adequately or "best."
- Some methods for quantitative comparison of results to observations are mentioned above, *i.e.*, mean, standard deviation, maximum error, mean squared error.
- At the end, we need a specific algorithm for quantifying how well a set of model results reproduces the selected observations.

Universidad Veracruzana

Simultaneous Results

- Of particular concern is how to calibrate several different kinds of model results at once.
- Example. If we want to calibrate a model to reproduce the number, size, and wealth of agents.
- How do we decide between a set of parameters that reproduces number and wealth well but not size, and a parameter set that reproduces size and wealth well but not the number of agents?
- Do we treat each kind of model result as equally important, or do we prioritize them?

Universidad Veracruzana

Simulation Experiments

- Now that the calibration criteria are carefully defined, we just need to run the model and find out what parameter values best meet them.
- It is tempting to just start trying parameter values and searching heuristically for good ones, but that approach is inefficient and very likely to miss good parameter combinations.
- The calibration experiment must execute the model many times, using combinations of values over the feasible range of all parameters.
- The results of this experiment will tell us what ranges of parameter values produce results that meet the calibration criteria.

Universidad Veracruzana

Setting up Experiments

1. Select values for the non-calibration parameters and input data, if any, that represent the conditions (the same time period, environment, etc.) under which calibration patterns were observed.
2. Identify the range of values considered feasible for each parameter, and decide how many values to use within that range, *i.e.*, the parameter space.
3. Then, setting up the experiment is very easy using the NetLogo's BehaviorSpace.
4. If your model takes a long time to execute so that runs are infeasible, you can start with fewer parameter combinations and then zoom in on the most promising area in a second experiment.

Universidad Veracruzana

Variability

- How to deal with variability in results?
- If a model is highly stochastic, the random variation in its results can make calibration seem more challenging.
- We cannot really calibrate a model to higher precision than justified by its stochasticity.
- Instead of estimating mean model results over a number of replicates, we recommend running more parameter combinations and analyzing them by plotting model results against parameter values.
- Use regression, or just your eye, to identify parameter values that produce results falling in the desired range.

Universidad Veracruzana

The Good and the Bad Cases

- Ideally, the final step of calibration would be to analyze the results of all the model runs in the calibration experiment and identify the parameter values that meet the calibration criteria, select the final values of the calibrated parameters, and document how closely the calibrated model fits the criteria.
- However, it is not at all unusual to learn from the calibration experiment that no combinations of parameter values cause the model to meet all the criteria.
- What should you do if you cannot meet all the calibration criteria at once?

Universidad Veracruzana

Dealing with Failure I

- First, find mistakes in everything from your submodels, software, and input data to the analysis of your calibration experiment.
- If everything is fine, then it is very likely that your model is too simple, in the wrong ways, to reproduce the observed patterns you chose as calibration criteria.
- You could consider going back to the theory development stage and seeing if you can improve the sub-models for agent behavior, or add new ones.
- But there are costs to adding complexity to your model, if it is not very clear what change needs to be made.
- Keep in mind the over fitting issue.

Universidad Veracruzana

Dealing with Failure II

- If you choose not to revise the model to make it fit more of the calibration criteria, simply document your calibration results and the extent to which the model does not meet some criteria under your best parameter values, and your decision not to revise the model.
- Then, when you use the model to solve problems, keep in mind which results are less certain as indicated by the calibration experiment.

The problem

- The Woodhoopoe model has few parameters, but some of them are particularly difficult to evaluate even with the extensive field studies conducted.
- The model is used exactly as before, but the sub-model for how subordinate adults decide each month whether to undertake a scouting foray is now:
  - If there are no older subordinate adults of the same sex in the territory, then do not scout.
  - Otherwise, decide whether to scout via a random Bernoulli trial with probability of scouting equal to the parameter `scout-prob`.
- Assume that this sub-model was tested in a theory development cycle and found to reproduce the characteristic patterns when `scout-prob` has a value around 0.3.

Universidad Veracruzana

Identify calibration parameters

- We assume that two parameters are good for calibration because they are particularly uncertain and expected to have strong effects on results:
  - `survival-prob` is the monthly survival probability for all birds.
  - `scout-prob`, the theory development cycle told us (we are assuming here) that it likely has a value around 0.3, but we do not have a certain estimate because scouting events are hard to detect in the field.

Universidad Veracruzana

Choose categorical versus best-fit calibration

- Instead of finding the "best" set of parameter values, we will look for values that produce results within an acceptable range of the calibration criteria.
- We choose categorical calibration because the data available for calibration are somewhat uncertain.

Universidad Veracruzana

Decide whether to use time-series calibration

- The Woodhoopoe model represents a quasi-stationary system with no inputs or processes that change over time.
- Only stochastic events cause the population to change over time.
- Hence, trying to calibrate its changes over time does not seem necessary or likely to succeed.
- Instead of time-series calibration, we will calibrate it against data averaged over time.

Universidad Veracruzana

Identify calibration criteria

- We will use three patterns as calibration criteria, as if they were observations from the field study:
  - The mean abundance criterion is that the long-term mean number of woodhoopoes (including sub-adults) is in the range of 115 to 135.
  - The variation criterion is that the standard deviation among years in the annual number of birds is in the range of 10 to 15 birds.
  - The vacancy criterion is that the average percentage of territories that lack one or both alphas is in the range of 15% to 30%.
- All the criteria are assumed to be from data collected in November (month 11) of each year of the field study (just before breeding).
- Results from the first two warm-up years of the simulation will be excluded.

Universidad Veracruzana

Design and conduct simulation experiments I

- Set up the BehaviorSpace to run an experiment that simulates many combinations of the calibration parameters and produces output corresponding to our calibration criteria.
- Just put statements like these in BehaviorSpace's "Vary variables as follows" field:

```
["scout-prob" [0 0.05 0.5]]
["survival-prob" [0.95 0.005 1.0]]
```

- To get output comparable to the calibration criteria, we need to write reporters in the BehaviorSpace dialog to produce output that we can analyze and compare to our criteria.
- Because the criteria include means and standard deviation among years, we must obtain output for each step.

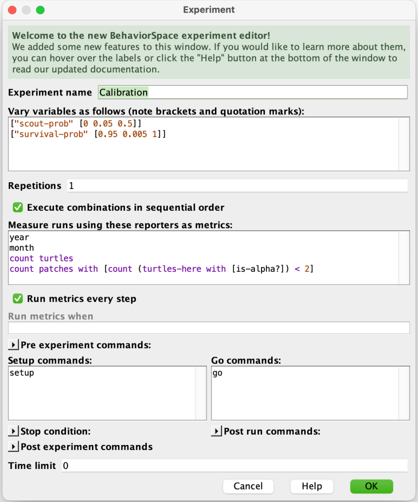
Universidad Veracruzana

Design and conduct simulation experiments II

- In addition to the number of woodhoopoes, we need the number of territories lacking one or both alphas.
- And because we need to exclude the first two years and examine only month 11, we need to output the year and month.

Universidad Veracruzana

The Experiment



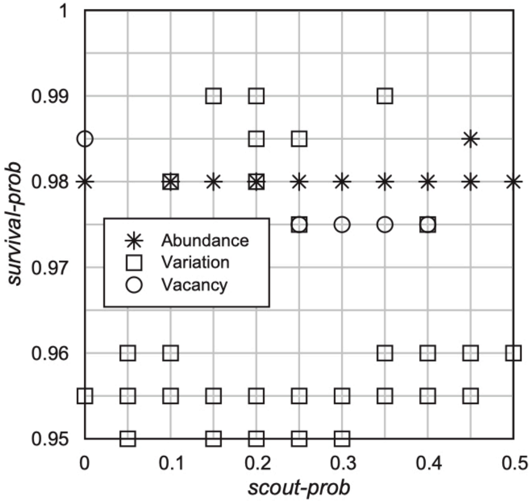
Universidad Veracruzana

Analyze Results

- We can calculate the fraction of territories lacking an alpha by simply dividing the values in the rightmost column by the total number of territories, 25.
- Then we need to filter the results and analyze only those from month 11 of years 3–22 (20 years after the 2 warm-up years).
- Tools: Pivot-Tables and Pivot-Charts.
- When we do this analysis in a spreadsheet, we find that there were no combinations of `scout-prob` and `survival-prob` where all three criteria were met. Does that mean the model is a failure?

Universidad Veracruzana

A Countour-Plot



Universidad Veracruzana

Observations

- The abundance criterion depended almost entirely on survival-prob (0.98).
- The vacancy criterion, though, was only met when survival-prob was 0.975 (and scout-prob was between 0.25 and 0.4).
- The variation criterion was met consistently when survival-prob was 0.96 or less, but also at some parameter combinations near where the abundance and vacancy criteria were met.
- Now the results do not look so bad: in the region around scout-prob = 0.25–0.4 and survival-prob = 0.975–0.98, there are locations close to where all three criteria were met.

Universidad Veracruzana

Referencias I

[1] SF Railsback and V Grimm. *Agent-Based and Individual-Based Modeling*. Second. Princeton, NJ, USA: Princeton University Press, 2019.

Universidad Veracruzana