Agent-Based Modeling and Simulation

Patterns for Model Structure

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

Universidad Veracruzana

Credits

- These slides are based on the book of Railsback and Grimm [4], chapter 17 and 18.
- Any difference with this source is my responsibility.
- This work is licensed under CC-BY-NC-SA 4.0
- To view a copy of this license, visit:

https://creativecommons.org/licenses/by-nc-sa/4.0/

Universidad Veracruzana

Structurally Realistic Models

- ABMs should be as simple as possible and use the most parsimonious representation that still captures key characteristics and behaviors of a real system.
- We refer to such models as structurally realistic.
- Aspects of the real system (entities, state variables and attributes, processes, spatial and temporal resolution and extent) should be included in the model only if they are absolutely essential for solving a particular problem.

Universidad Veracruzana

Pattern-Oriented Modeling

- Pattern-Oriented Modeling (POM) is the use of patterns observed in the real system as additional information to make ABMs structurally realistic and, therefore, more general, useful, scientific, and accurate.
- Example. Many population models explain the pattern of strong cycles in the abundance of small mammals in boreal regions, *e.g.*, lemmings. Some of these models are not structurally realistic but instead wrong about the mechanisms causing population cycles, *i.e.*, they reproduce the cyclic pattern for the wrong reasons!

Universidad Veracruzana

What is a pattern?

- A pattern is anything beyond random variation. We can think of patterns as regularities, signals, or, as they are sometimes called in economics, stylized facts.
- Example. Light emitted by an excited atom includes only a few characteristic wavelengths in the spectrum. Atoms of different elements can be distinguished by their spectra. Thus, the spectra are beyond random variation: they call for an explanation because they indirectly reflect the structure of atoms.

Universidad Veracruzana

Single and Multiple Patterns

- For most systems, however, one single pattern is not enough to decode the internal organization.
- Multiple patterns, or filters, are needed.
- Example. When you need to meet someone you do not know at the airport, a single pattern, *e.g.*, knowing that he is male, is not enough to identify the right person. But if you also know that the person's age is about 30, and that he will wear a green T-shirt and carry a blue suitcase, you probably will identify him right away.

Universidad Veracruzana

Observations

- Things to note about this set of patterns:
  - It is small –only four patterns;
  - The patterns are simple or weak, each captured in just a few words and none strongly descriptive;
  - The patterns are qualitative, not quantitative;
  - The patterns are about different things –sex, age, clothing, and baggage; and
  - It is relevant specifically to this problem –the patterns are useful because they can be detected at a distance, but you would use a different set to describe the same person for a different problem, *e.g.*, deciding whether he should marry your sister.
- They can be as powerful a filter as a strong pattern, *e.g.*, a photograph of the person, and are often easier to obtain.

Universidad Veracruzana

Characteristic Patterns

- Patterns that we believe, from what we know about the system, to be caused by the same mechanisms and processes that drive the system's behavior relevant to the problem we model.
- If your model do not reproduce them, you should not trust it to solve your problem.
- Example. An European beech forest model that reproduced one key pattern of real forests (a horizontal mosaic of developmental stages), but was not trusted by beech forest experts because it was not capable of reproducing additional patterns (in vertical structure) that they considered essential.

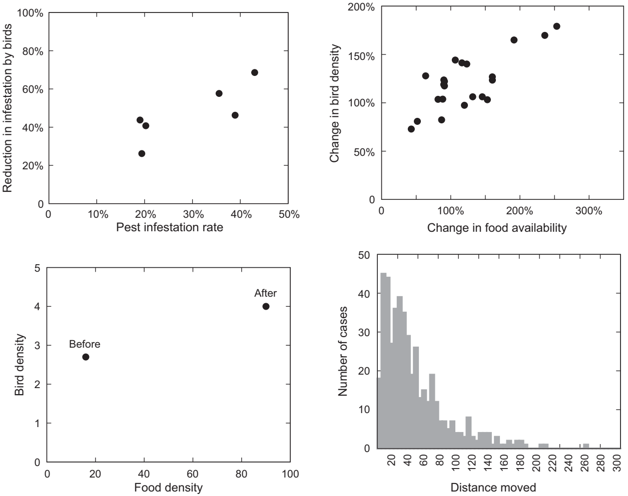
Universidad Veracruzana

Qualitative of Quantitative

- POM includes the step of testing whether a model reproduces specific patterns, so we need to define criteria for when a pattern is matched.
- Do we need to define the patterns numerically and designate some statistical tests for whether model results match them?
- At least until the final stages of modeling, it is important not to be rigidly quantitative; otherwise, it is too easy to get trapped in a bog of statistical inference before understanding anything.
- Example. Watson and Crick found their model of DNA's structure using qualitative criteria first; only later did they turn to quantification [5].
- Proof usually comes first, quantification second.

Universidad Veracruzana

Example. Pest Control in Coffe farms



Universidad Veracruzana

Observations

- The reduction in pest infestation (percentage of coffee berries infested by insects) is higher when the infestation rate is higher;
- Over a season at specific sites, the change in bird density is positively related to the availability of nonpest food;
- Bird density increases rapidly after a sudden insect outbreak that increases the density of nonpest food;
- The distribution of distances that birds move within 1 hour follows a log-normal distribution, with many short-medium distance moves and few, but not zero, long moves.

Universidad Veracruzana

Role of patterns

- The basic idea of POM is to use multiple patterns to design and analyze models.
- Each pattern serves as a filter of possible explanations so that after using, say, three to five patterns, we are more likely to have achieved structural realism, *i.e.*, an unrealistic model might pass one filter, but probably not also a second and third.
- POM thus uses multi criteria assessment of models using multiple patterns that have been observed in the real system, often at different scales and at both the agent and system levels.

Universidad Veracruzana

Uses of POM

- Using POM for agent-based modeling includes:
  - Using patterns to design a model's structure;
  - Developing and test theory for agent behavior; and
  - Tune parameter values.
- In this session we focus on the first use of POM.

Universidad Veracruzana

The problem I

- The first very important decision in designing a model is selecting the set of entities and their state variables and attributes that represent the system.
- Then we also know what processes need to be in the model, *i.e.*, those that are important for making the state variables change.
- Although the purpose of a model is a guide for such decisions, it can provide too narrow a point of view, leading to a model that is too simple in structure to be tested thoroughly.

Universidad Veracruzana

The problem II

- Example. For predicting the growth of the human population, the simplest model could be an exponential growth equation with the population size as its only state variable and one parameter fit to historic population data. But then how would we know whether the model is a sufficiently good representation to project the population for 20 years, or 100 years, into the future?
- Including too few entities and state variables causes models to be under determined, *i.e.*, they do not, due to their simplicity, provide enough kinds of results to be tested adequately. They are too poor in structure and mechanisms [1].

Universidad Veracruzana

Solution

- We'll use observed patterns to design model structures that includes enough to make the model sufficiently realistic and testable, but not unnecessarily complex.
- Example. Knowing a lot about the age structure of human populations, how many people there are of each age, in different regions. A model that is capable of reproducing historical patterns in both total population size and regional age structures is much more likely to include the key processes of demography than a model focusing only on global population size.

Universidad Veracruzana

Objectives

- Understanding the four steps in using Pattern-Oriented-Modeling (POM) to design a model's structure, and the examples of this process that are presented here.
- Be able to structure an ABM from:

  1. Knowledge of the system and problem that the model addresses and
  2. Observed patterns that characterize the system's processes relevant to the problem.

Universidad Veracruzana

Initial Formulation of the Model

- Start by formulating your model using the ODD protocol, with the purpose as the only filter for designing the model structure.
- Include the entities, state variables, and processes that seem the minimum necessary to solve the problem you model, given your current understanding or conceptual model of the system.
- At this stage the formulation should seem too simple to possibly be useful.

Universidad Veracruzana

Identifiy Observed Patterns I

- Identify a set of observed patterns that you believe characterize your system relative to the problem you model.
- Spend the time necessary with the literature and with experts to really understand what is known about the system and how its processes are believed to give rise to the patterns, especially if you do not know the system well already.
- Limit yourself to a manageable set of patterns, typically two to six, most often three or four.
- Often, some patterns are closely linked and not independent of each other.

Universidad Veracruzana

Identifiy Observed Patterns II

- Try to find a diversity of patterns, such as responses to several different kinds of change.
- The choice of patterns is experimental at first because we cannot know in advance which patterns contain the information we need to design a model.
- Try to rank the patterns by how important they are in characterizing the system and problem.
- Document these patterns in the Purpose and Patterns section of your ODD model description.

Universidad Veracruzana

Define Pattern-Matching Criteria

- How will you decide whether the model does or does not reproduce each pattern?
- Start with qualitative criteria such as visual pattern matching and trends in statistical summary output, though quantitative criteria may be needed in later steps of POM.
- We discuss pattern-matching more in the two following sessions; for now, the only concern is having some idea of what outputs you will use to evaluate whether the model reproduces your characteristic patterns.

Universidad Veracruzana

Review the Draft Model Formulation

- Determine what additional things need to be in the model to make it possible for your characteristic patterns to emerge from it.
- Are any new entities or state variables needed so that the patterns could emerge?
- Are any of the patterns believed to be driven by processes that are not already in the model?
- Are new variables or outputs needed to observe whether the patterns are reproduced?
- Whatever you identify, typically only a few variables and processes, should be added to the model formulation.

Universidad Veracruzana

Expected results

- Now you are ready to advance in a second tour through the modeling cycle.
- Revise the draft formulation to include the changes identified in these steps, then program the model and test the software thoroughly.
- When you have done this, you will be ready for the next stage of POM: testing your ABM and its theory for individual adaptive behavior, the subject of the next session.

Universidad Veracruzana

A Story of Two Models

- These models of spatio-temporal dynamics shows the use of characteristic patterns to design ABMs and how their choice can affect a model's structure and, consequently, its believability and success.
- After the most recent ice age and before large-scale agriculture, much of central Europe was covered by forests dominated by beech trees.
- These forests consisted of a mosaic of patches of large, old trees with a closed canopy and virtually no understory, and patches with gaps in the canopy and a mixture of young and old trees.
- Their purpose is to understand the mechanisms causing this mosaic pattern and, ultimately, to decide how big and old forest preserves must be to resemble the ancient forest.

Universidad Veracruzana

Model 1

- The model of Wissel [6] is a simple grid-based model.
- Its design focused on reproducing a single pattern observed in natural beech forests: the mosaic of patches in the stages described before.
- The patch types are believed to follow each other in a development cycle: a canopy that is closed opens when old trees die due to storms or age, which lets younger trees establish and grow in a mixture of ages and heights, until the canopy is closed again and most of the small trees die because they no longer get sufficient light.

Universidad Veracruzana

Structural Design I

- The model needed to include only patches with state variables only for which developmental stage the patch is in and how long it has been in that stage.
- Trees grow slowly, so the time step is 10 years.
- The model's main process is the progression of each cell through the cycle of developmental stages.
- This cycle assumes that cells progress from early colonizing species such as birch, to a mixture of later colonizers, followed by stages dominated by young and then older beech trees.
- In the final stage of the cycle, a cell contains a single large beech tree, until this tree dies.

Universidad Veracruzana

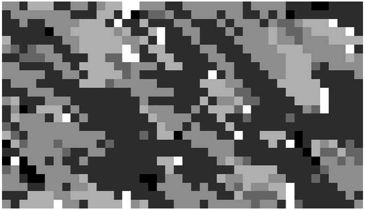
Structural Design II

- The model's second process is an effect of one cell on its neighboring cells: beech trees can be killed by too much sunlight on their bark, so when a cell changes from the large beech tree state to open canopy (the big tree dies), the cells to the north (and, to a lesser degree, east and west) are more likely to have their trees die soon.

Universidad Veracruzana

Graphically

- The output:



- The shading represents the time since a cell was empty: cells range from white on the first time step after becoming empty to black at the maximum age of a cell.

Universidad Veracruzana

Controversy

- Landscape ecologists, who focus on such patterns, were quite happy with this model.
- However, beech forest scientists were not convinced that the model produced these patterns for the right reasons.
- They questioned several model assumptions and, in particular, believed that the developmental stages were not automatic but instead emerged from processes that occur mainly in the vertical dimension and produce important vertical patterns.
- Vertical competition for light was believed to be the most important structuring process: if the forest canopy is closed, too little light reaches the understory, leading to high mortality of smaller trees.

Universidad Veracruzana

Additional Patterns

- Forest scientists saw an additional set of patterns: the vertical characteristics of the different developmental stages.
- When large trees die, their patch cycles through stages with only small seedlings, then taller saplings, and finally tall trees again.
- Further, the forest scientists believed they knew why most big trees die: they are typically toppled by high wind during storms.

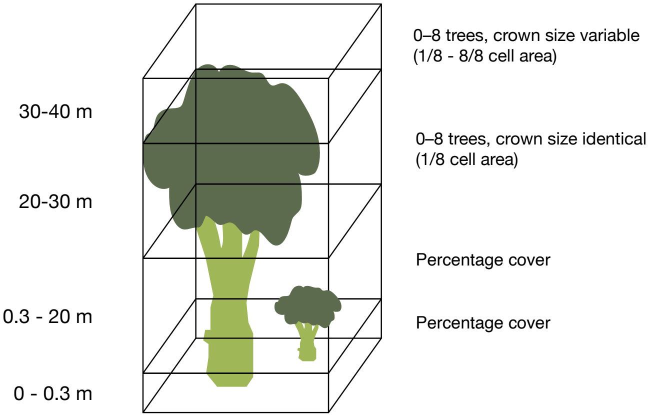
Universidad Veracruzana

Model 2: BEFORE

- Two kinds of entities: vertically compartmented grid cells and individual large beech trees. Trees are tracked as individuals only after they grow out of the juvenile stage.
- It is simpler than the original model in one way: it ignores tree species other than beech.
- Horizontally, grid cells are 14.29 meter squares (the area covered by one very large beech). Vertically, cells have four compartments that each represent a height class of beech.

Universidad Veracruzana

Graphically



Universidad Veracruzana

State Variables I

- Height compartments are separate entities: Upper canopy, Lower canopy, Juveniles, and Seedlings.
- Upper and lower canopy have a list of the trees present in the compartment. Zero to eight trees can be present in each of these compartments of each cell.
- These two compartments each have a state variable (F3 and F4) for the percentage of the cell area occupied by trees in the compartment.
- Juveniles and seedlings compartments each have one variable, (F2 and F1), for the percentage of cell area covered by them.
- The juvenile compartment also has a state variable for the number of previous time steps in which F2 was greater than zero.

Universidad Veracruzana

State Variables II

- Upper canopy trees have variables for age and their crown area measured in octaves.
- Lower canopy trees have variables for age, crown area, and time spent in the compartment.
- Time steps represent 15 years.
- The spatial extent is 54 $\times$ 54 grid cells, representing 60 hectares.

Universidad Veracruzana

Processes

- All processes occur within grid cells and among neighboring cells.
- The establishment, growth, and mortality of trees all depend on vertical competition for light, which is determined by the variables F2, F3, and F4, representing how much of the cell area is covered by trees.
- Toppling of big trees by windstorms is a stochastic process. Windtoppled trees can hit and destroy trees in neighboring cells.
- The opening of gaps in the canopy affects neighboring cells by making their trees more vulnerable to future windstorms, but also by allowing slanted light to reach their lower compartments and promote growth.

Universidad Veracruzana

Observations

- Since BEFORE was designed using more patterns, and more diverse patterns, it is more complex than the original model, although still conceptually simple.
- It successfully reproduced emergent patterns, including:
  - The observed mosaic of developmental stages over space,
  - The cycle of developmental of stages over time in a cell, and
  - System-level properties such as the proportion of the forest in the final stage (upper canopy trees) and the cyclic variation of this proportion.
- The model was thus successfully verified.

Universidad Veracruzana

Secondary predictions I

- BEFORE is structurally rich enough to produce output that can be scanned for new patterns that were not known, or considered, when the model was formulated, tested, and calibrated.
- We call such new patterns secondary or independent predictions.
- If such predictions are confirmed by new empirical evidence (literature, measurements, or experiments), the model is said validated instead of just verified.
- Even if we use multiple observed patterns for verifying a model, there is still a risk that we tweak it to reproduce the right patterns for the wrong reasons, *e.g.*, by calibration.

Universidad Veracruzana

Secondary predictions II

- So, we are searching for evidence of the model's structural realism that was not used during model development and verification.
- Example. BEFORE was analyzed for patterns in the age distribution of the canopy trees and the spatial distribution of very old and very large canopy individuals (giants), but only after BEFORE was designed and calibrated.
- Example. BEFORE was modified to track how much dead wood (which provides habitat for many species) the model forest produces.

Universidad Veracruzana

Ideas

- Using a variety of observed patterns to design an ABM's structure produces a model that is more structurally realistic, testable, and believable.
- Could we not apply this lesson in reverse and evaluate how believable existing models are by how many patterns they were designed to explain?

The problem

- Heine, Meyer, and Strangfeld [3] discuss patterns in a management problem of large businesses: the decision by central management of how much money to invest in the company's different divisions.
- Groves [2] invented a mechanism of distributing investment among divisions in a way that compels managers to predict profits honestly.
- However, this mechanism for eliciting honest management information can be undermined by collusion among division managers.

Universidad Veracruzana

Patterns

- The authors identified the following general patterns or stylized facts under which collusion is favored:

  1. When the number of potential colluders (number of divisions in a company) is small,
  2. Under long time horizons,
  3. In uncomplicated settings (low dynamics, low complexity, high transparency, etc.),
  4. When the potential benefits of collusion are high,
  5. When group composition is stable, and
  6. When there are effective ways for colluders to enforce their agreements.

Universidad Veracruzana

Results

- Four models are compared, two based on game theory and two in simulation.
- Only the second simulation model second had much credibility for explaining collusion in the management accounting problem; only that model considered enough observed patterns to leave “little room for arbitrary model design or questionable parameter calibration”

Universidad Veracruzana

Referencias I

[1] DL DeAngelis and WM Mooij. "In praise of mechanistically-rich models". Niet gedefinieerd. In: *Models in Ecosystem Science*. Ed. by DC Canham, JJ Cole, and WK Laurenroth. Princeton, NJ, USA: Princeton University Press, 2003, pp. 63–82.

[2] T Groves. "Incentives in Teams". In: *Econometrica* 41.4 (1973), pp. 617–631.

[3] BO Heine, M Meyer, and O Strangfeld. "Stylised Facts and the Contribution of Simulation to the Economic Analysis of Budgeting". In: *Journal of Artificial Societies and Social Simulation* 8.4 (2005), pp. 1–34.

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

[5] JD Watson. *The double helix: A personal account of the Discovery of the Structure of DNA*. New York, NY, USA: Atheneum, 1968.

[6] C Wissel. "Modelling the mosaic cycle of a Middle European beech forest". In: *Ecological Modelling* 63.1–4 (Sept. 1992), 29–43. ISSN: 0304-3800. URL: http://dx.doi.org/10.1016/0304-3800(92)90060-R.

Universidad Veracruzana