UNIVERSITÄT LEIPZIG

Toolbox CSS – Agent-based Models II

NSG SR 423, 21/01/2025

Felix Lennert, M.Sc.

ABMS – WHERE WERE WE…

"Agent-based models are computer programs in which artificial agents interact based on a set of rules and within an environment specified by the researcher" (Bruch & Atwell 2015, p. 187)

OUTLINE

- How to report ABMs
  - ODD protocol
- Empirically calibrated ABMs – a primer
  - Idea
  - In practice

BUILDING ABMS FROM SCRATCH – THE ODD PROTOCOL (GRIMM ET AL. 2006, 2010, 2020)

- A standardized way of reporting ABMs
- Easy for the reader/reviewer
- But also very helpful as a schema to construct ABMs from scratch

BUILDING ABMS FROM SCRATCH – THE ODD PROTOCOL (GRIMM ET AL. 2006, 2010, 2020)

Overview	O	1. Purpose and patterns
		2. Entities, state variables and scales
		3. Process overview and scheduling *Submodel A* *Submodel B* …
	D	4. Design concepts
Details	D	5. Initialization
		6. Input data
		7. Submodels *Submodel A (Details)* *Submodel B (Details)* …

Basic principles
Emergence
Adaptation
Objectives
Learning
Prediction
Sensing
Interaction
Stochasticity
Collectives
Observation

Overview	
	1. Purpose and patterns
	2. Entities, state variables and scales
	3. Process overview and scheduling

Purpose and patterns:
- What is the model supposed to show – which patterns or phenomena

Entities, state variables, and scales
- Description of model entities (agents, environment)
- Definition of state variables and attributes
- Specification of temporal and spatial scales

Process overview and scheduling (make a flow chart or table, e.g., Baldassarri & Bearman (2007), Table 1)
- List of model processes
- Order of execution
- Time stepping

D	4. Design concepts

Basic principles
Emergence
Adaptation
Objectives
Learning
Prediction
Sensing
Interaction
Stochasticity
Collectives
Observation

basic principles – are there existing, similar models? What are the differences?

emergence – how do results come about

adaptation – how do actors adapt

sensing – how much do actors know about the state of other agents?

learning – does new knowledge lead to new decisions in actors?

prediction – are actors trying to predict the future?

interactions – how do actors interact?

stochasticity – are there any “random numbers” involved?

collectives – do we see groups of actors?

observation – how do you collect and summarize information?

Details	D	**5. Initialization**
		6. Input data
		7. Submodels *Submodel A (Details)* *Submodel B (Details) …*

Initialization:

- How's the world set up
- How are the actors initialized
- How are the state variables set up?

Input data:

- Particular input parameters (e.g., empirical calibration, GIS data, etc.)

Submodels:

- Report all processes (formulas, parameters and how they change, algorithms),
- Justify them
- Robustness checks – how do parameters change behavior

=> Align this with "Process overview and scheduling"

SO FAR: “ABSTRACT” ABMS

“some micro-level behavior is known or strongly assumed and simulation explores its aggregate consequences” (Bruch & Atwell, 2015, p. 192)

=> example: Schelling model – highly unrealistic

But: “the goal is not to reproduce existing patterns or even to anchor agents’ behavior, characteristics, or environment in empirical knowledge … the models … develop new ways of thinking about a problem”

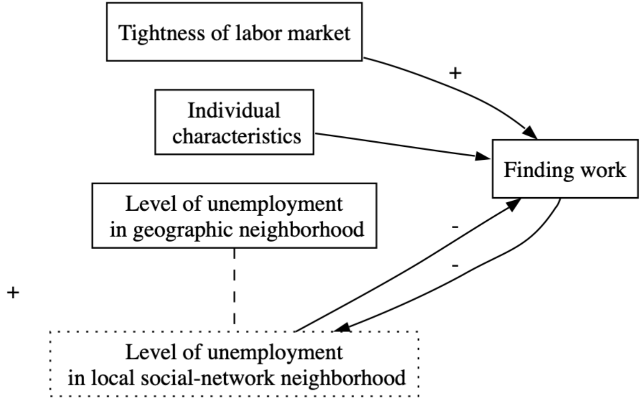
=> KISS – keep it simple, stupid

"LOW-DIMENSIONAL REALISM" ABMS

"aimed at exploring the implications of empirical research or testing the assumptions of formal theories" (Bruch & Atwell, 2015, p. 193)

=> example: Hedström & Åberg (2005): unemployment dynamics

- Incorporate social and demographic characteristics of young adults in Stockholm
- Vary networks of young adults
  => If many people in your environment are unemployed, you don't feel pressured to find a new job



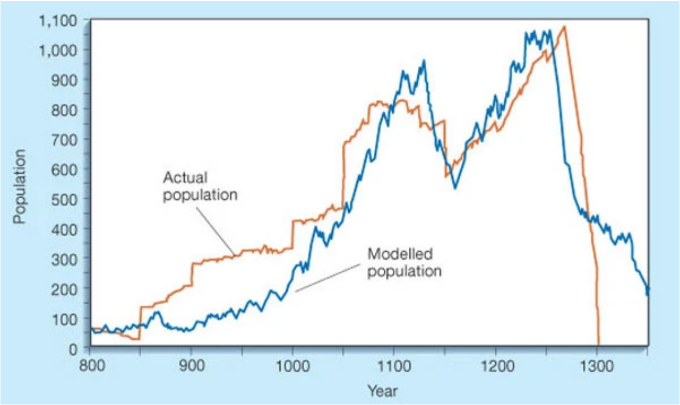
Goal: "not to reproduce empirical patterns or incorporate all aspects of reality so much as to understand the implications for social dynamics of one or more empirical observations or stylized facts" (Bruch & Atwell, 2015, p. 193)

=> KIDS – keep it descriptive, stupid

"HIGH-DIMENSIONAL REALISM" ABMS

Example: Artificial Anasazi project (Dean et al. 2000)

- "How can the complex dynamics of human societies — such as population rise and fall, and movement — be explained? Combining masses of data with computer modeling is a fresh way forward." (Diamond 2002)
- Modeling the rise and fall of the Anasazi culture in AZ
  - simulated virtual Anasazi farmers into the valley at AD 800
  - data on rainfall, groundwater, soil types, crop yields and household behavior
  - feeding them each year with the calculated corn crop
  - letting them bear and feed children, grow old, move house sites, and send off grown children to build new houses => according to rules observed for recent corn-growing societies of Pueblo Indians descended from the Anasazi
  - Finally: spatial distribution of farms, population numbers, compared these to empirical outcomes



Fairly accurate estimates, good for prediction

WHICH ONE TO CHOOSE?

=> Depends on the goal

- Testing a theoretical mechanism: keep it as simple as possible to observe your outcome – e.g., Schelling model – looking for broader patterns
- Accurate prediction (e.g., modeling in epidemiology): bring in as much as possible – point estimates matter!
- Showcase a mechanism and compare it to empirical outcome – middle ground, bring in as much as needed, as little as possible – point estimates don't matter, but general patterns do

EMPIRICAL CALIBRATION

- Crucial parameters: *empirical realism* and *level of complexity*
  => Independent of each other

- Empirical realism: behavior of people grounded in social psychology (for instance)
- Complexity: number of parameters/characteristics (e.g., population data on age distributions, geographical features, rainfall, etc.)

IN PRACTICE

A*gents* have attributes (e.g., skin color, gender, education, income, etc.)
=> Can be incorporated arbitrarily (based on some probability distribution)
=> Can be incorporated based on some observed distribution (e.g., census data)
=> Also: network connections between agents

- Common problem: we only observe aggregate values
- Solution: *iterative proportional fitting* (this is of course not perfect if things are unequally distributed, e.g., gender and education)

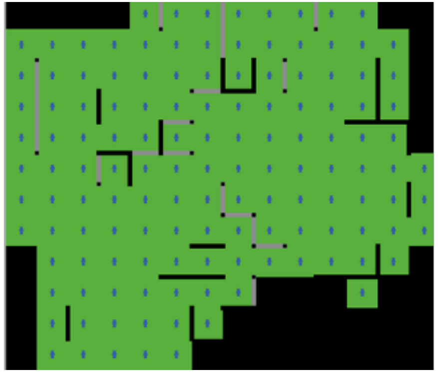
Car Ownershi	Household Size: Size 1	Size 2	Size 3	Size 4	Size 5	Size 6 +	Row total
0 Cars							146600
1 Car							405209
2 Cars							162646
3+ Cars							18338
Column total	368447	211351	77445	52208	15785	7557	

Car Ownershi	Household Size: Size 1	Size 2	Size 3	Size 4	Size 5	Size 6 +	Row total	Actual	Ratio
0 Cars	73710.2	42282.1	15493.4	10444.5	3157.9	1511.8	146600	146600	1.000
1 Car	203738.4	116869.7	42824.4	28869.2	8728.6	4178.8	405209	405209	1.000
2 Cars	81778.1	46910.1	17189.2	11587.8	3503.5	1677.3	162646	162646	1.000
3+ Cars	9220.3	5289.0	1938.0	1306.5	395.0	189.1	18338	18338	1.000
Column total	368447	211351	77445	52208	15785	7557			
Actual	368447	211351	77445	52208	15785	7557			
Ratio	1.000	1.000	1.000	1.000	1.000	1.000			

IN PRACTICE (BRUCH & ATWELL 2015)

Agents inhabit an environment
=> Can be modeled with low empirical realism – cellular automaton
=> Can be modeled realistically – ideally: GIS data with borders, roads, etc.

… or whatever this is (Hägerstrand 1965):



IN PRACTICE

Agents have certain behaviors/empirical preferences
=> Typically: gather information from environment, assess these information based on some criteria/ranking system, make decisions
=> All this needs to be specified – depends on model purpose:

- If abstract: needs to be aligned with theory
- If higher level of realism: needs to be empirically defensible (e.g., Hedström & Åberg 2005)

- Useful strategy: assume some sort of statistical model of change or choice
  - For attribute change: discrete-time event history model (Allison 1982)
  - For choices (e.g., interlocutor): discrete choice model
  - Independent variables: revealed preferences (observational data); stated preferences (survey); characteristics; values; etc.
    => Multiple data sources can be combined

AFTER THE SIMULATION

- In theory you can look at each agent at every point in time/step (micro level)
- In practice you should look at macro descriptors (e.g., segregation measures, number of dead agents, etc.)
- There is usually some **uncertainty** and **variability** (requiring sensitivity analyses)
  - *Input uncertainty*: different input specifications may lead to different outcomes
  - *Model uncertainty*: model architecture might have some impact as well (e.g., number of agents

=> Solution: try different specifications, compare outcomes

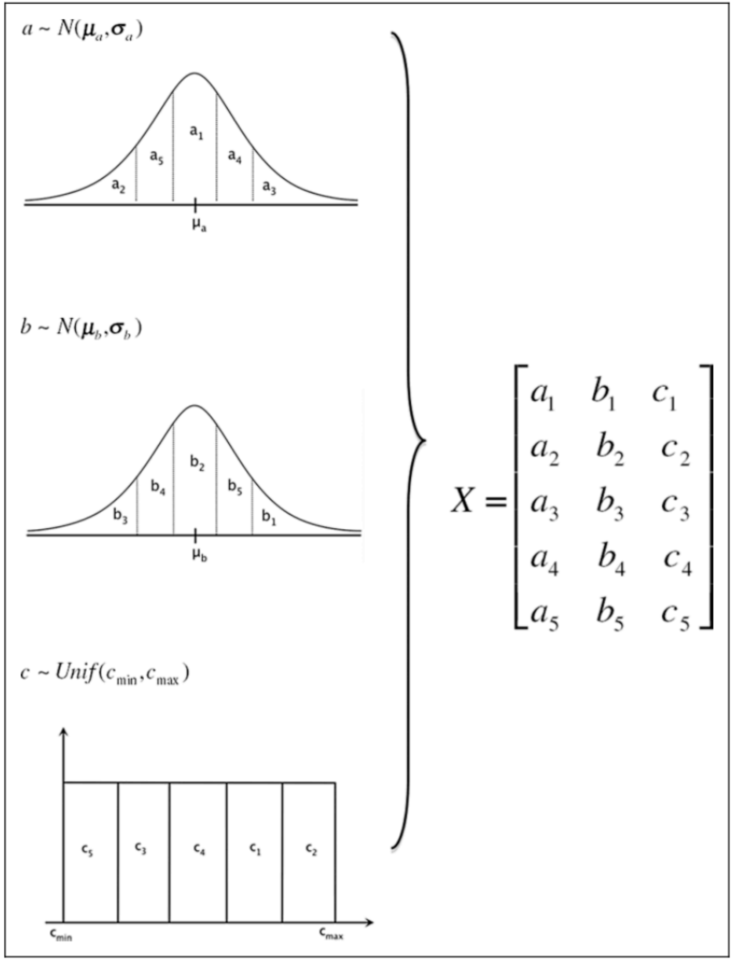
  - *Stochastic variability*: variation across runs with same specifications
    => differences can add up due to path dependencies

=> Solution: multiple runs with same specs, aggregate results

AFTER THE SIMULATION

For sensitivity analyses, you can just do a grid of different parameters and run through them
=> However, takes a while

Solution: Latin Hypercube Sampling:
1) specify distributions of parameters
2) slice them up into n random chunks (here: n=5)
3) draw random numbers from chunks
4) run analysis using the different parameters



VALIDATION

- Internal validity – usually high
  => researcher maintains full control, coded model after all
- External validity – needs to be checked
  - Not alway possible…
    - Agent population needs a certain real-world equivalent
    - Time steps need a certain real-world equivalent
    - Locations need a real-world equivalent
  - Measures are compared on different levels
    - Macro: final result or development over time (if you have longitudinal data)
    - Micro: Agent behavior

=> If you have an abstract ABM: can you derive hypotheses from the model that you can test in the real world?

NEXT WEEK

- Deadline for sending me your presentations: **Friday, January 31, 18.00**
- Also: regarding presentations – keep them shorter, maybe 5-7 minutes of you talking, 5 minutes feedback from opponent, 3-5 minutes feedback from audience
  => I'll cut you off, ruthlessly

UNIVERSITÄT
LEIPZIG

MERCI

Felix Lennert
Institut für Soziologie

felix.lennert@uni-leipzig.de
www.uni-leipzig.de