Overview of the Modeling Process 2: Formulation & Beyond

Nathaniel Osgood

Using Modeling to Prepare for Changing Healthcare Needs Duke-NUS April 16, 2014

ABM Modeling Process Overview



Model Formulation

- Model formulation elaborates on problem mapping to yield a fully specified, quantitative model
- Key missing ingredients: Specifying unambiguous specification for
 - Statechart transitions
 - Flows (in terms of other variables)
 - Observer processes
 - Intermediate variables
 - Parameter values

Model Specification Mechanisms

Stock & Flow Models: "Hedgehog Knowledge"

- Small modeling vocabulary
- Power lies in combination of a few elements
- Analysis conducted predominantly in terms of elements of model vocabulary

Agent-Based Modeling: "Fox Knowledge"

- Large modeling vocabulary
- Different subsets of vocabulary used for different models
- Power in flexibility & combination of elements
- Variety in analysis focus

ABMs: Larger Model Vocabulary & Needs

- Events
- Multiple mechanisms for describing dynamics
 - State diagrams
 - Stock and flow
 - Custom update code
- Inter-Agent communication (sending & receiving)
- Multiple types of transitions
- Diverse types of agents
- Spatial & topological connectivity & patterning

- Subtyping
- Mobility & movement
- Graphical interfaces
- Data output mechanisms
- Stochastics complicated
 - Scenario result interpretation
 - Calibration
 - Sensitivity analysis
- Synchronous & asynchronous distinction, concurrency

Process Interaction & Scheduling

- In addition to specifying the processes in isolation, try to describe process interaction e.g.
 - A transmission process is not triggered until a person is sexually active
 - All reporting takes place at the very end of the day, and is done before resetting reporting counters
 - All agents first note the status of the agents around them, and only then perform updates to location
- Ask yourself on what other processes a given process depends

Concurrency

- Two or more processes may be operating concurrently ("in parallel")
 - e.g.: Operation of different agents, agents & reporting processes, graphical interface & model

Dependencies: Synchronous vs. Asynchronous

- Suppose process A depends on information produced by process B
 - e.g. depends on knowing something produced via B
- Synchronous processes: Applied sequentially, so that A must wait for B to proceed (e.g. A calls B)
- Asynchronous processes: No "blocking" (waiting) by A for B (e.g. B sends a message to A)
 - In agent-based modeling, most interactions between agents are considered *asynchronous =>* inter-agent communication is accomplished via asynch. messaging

ODD: 3 Broad Components

- Overview: model goals & high level scope & design
- Design concepts: Different aspects of design being considered
- Remaining elements

ODD Design Concepts to Consciously Consider

- Origin & character of basic principles underlying model
- Emergence: To what degree are results pre-programmed vs. arising naturally out of a myriad of interactions
- Adaptation: How does system evolution lead to entity behavior change?
- Sensing: What information do entities retrieve from world?
- Objectives: Any goal seeking behavior? How interacts w/state?
- Learning: How does experience drive change in strategies?
- Prediction: How do entities anticipate the future?
- Interaction: How do entities interact directly & indirectly?
- Stochastics: Character of & motivation for stochastic effects
- Observation: What information & associated processes are required for operational use or for testing & confidence bldg

Sensing

- Information sensed from other agents & environments is key to adaptation & decisions
- Need to consider what is sensed
- May want to capture fact that entity perception is
 - Localized (e.g. risk perception, cf decision making with driver's view of road compared to with perfect knowledge of traffic flows across city)
 - Error prone
 - Delayed
 - This can fundamentally alter dynamics: e.g.
 - Instability: Fragility of "Tit for Tat" to misunderstandings
 - Negative feedback: Sensing to correct driving path

Emergence

- To what degree are the results directly captured by assumptions? (i.e. to what degree are we presupposing what we are trying to demonstrate?)
- One ABM viewpoint: Until we can robustly generate a phenomenon, we don't really understand it
- To what degree to results emerge from complex interaction of other factors where the behavior of interest is never itself described in any way
 - This is ideally what is sought
 - it allows more of a real explanation
 - Permits greater generality (anticipating system behavior under unobserved situations)
 - e.g. waves of infection in spatial SIR model
 - In CWD Model: Clustering of prions along
 - the lakeshore margin
 - High traffic cooridors

Emergent Behavior: Spatial/Geographic



A Multi-Level (Dynamic) Model



Adaptation

- How does agent *behavior* exhibited depend on the
 - Local or global environment
 - Surrounding agents
- To what degree is agent behavior fixed based on predefined rules (just playing out to understand collective effect of rules themselves) vs. potential for emergence associated with inter-agent or agentenvironment *behavioral* interaction, which often leads to correspondingly richer emergent behavior
 - Note that can still have inter-agent emergence without adaptation -- e.g. in an infection spread model. But the presence of adaptation means that the feedbacks and emergent phenomena can be that much richer

How Does Behavior Depend on Context?

- We have great flexibility in representing agent rules
- Some agents may be consciously objective seeking
- Just reproducing statistical patterns (likelihood changes in tobacco use over time)
 - Limited generality under counter-factuals
- Examples of ways might depend on context
 - Behavior change due to risk perception
 - Moving to a new neighborhood or hunting/gathering ground
 - Remembering insults and changing strategies (e.g. to defect) with respect to a neighbor in a connection matrix
 - By acquiring new memes or information from a neighbor

Incorporating Observed Patterns: 3 Ways

- Building patterns directly into model (likelihood of state transitions, mixing matrix per observations)
 - e.g. Fraction of time spent in different states (foraging, new lake margin, near grain bins)
 - E.g. fraction of time spends with different groups
- Building functional dependence of actions on external conditions into the model
 - E.g. mixing matrix as a function of a preference matrix and current population demographics
- Calibrating or validating to patterns
 - Making patterns emerge from lower-level "mechanics"/ "physics" of model
 - e.g. Contacts (or contact networks) emerge from myriad closeproximity spatial interactions between mobile individuals

Observed Patterns as Emergent

- Ideally, we seek to make patterns emerge from lower-level "mechanics" / "physics" of model
 - e.g. seasonal herd size emerging naturally from grouping rules in CWD model
- With *adaptation*, particularly focusing on dependence of behavioral patterns of an individual on context
 - How do varying circumstances change agent behavior?

Example of Observed Patterns as Emerging from Low-Level Interactions

- Lower food availability => Higher amount of time spent searching for food
- Higher prevalence of Gonorrhea among acquaintances => greater adherence to safer sex practices
- Higher reports of H1N1 infection or vaccination among social contacts => higher chance of getting vaccinated
- Higher risks from diabetes over age as emerging naturally from cumulative damage by glycosylation, etc.
- Greater smoking-related health complaints & sickness in peers with age => Greater likelihood of quitting with age
- Progression of substance abuse caused by underlying organic processes
- Longer infectious period, greater infection severity (peak viremia level), greater transmissibility for individuals with impaired immune functioning emerging from immune repr.
- Higher temperature => greater water seeking

One Kind of Adaptation: Objective Seeking Behavior

- Here, an entity's behavior will depend on trying to maximize some satisfaction criteria
 - Examples of measures: Profit, Utility
 - Example application: Vehicle simulators using where driving behavior depends on consideration of perceived tradeoffs (\$, time, familiarity, etc.) of different routes
- How does this vary based on agent's state (e.g. access to resources) or environments
- Bounded rationality: For individuals, strong literature suggests that many decisions are based instead on heuristics

Learning: Changing Adaptive Behavioral Rules Based on Experience

- ABMs can support arbitrarily rich learning that may change adaptive behavior
 - Learning from experience in particular healthcare facilities
 - Trust of different parties based on
 - Direct: Treatment received
 - Indirect: Consistency of observations with claims of other party
- In some cases, this is performed using genetic programming (rules mutate and evolve)
- As a longitudinal phenomenon -- one that involves history -- support of learning & memory is a key advantage offered by ABM

Interaction

- Interaction among entities
 - Agent-agent
 - Agent-environment
- Forms
 - Direct: Agents directly interact with neighbors (e.g. via needle sharing or sexual contact)
 - Indirect: e.g. Via shared resource (depletion of vegetation for browsing by other deer, deposit of droplets with shedded pathogen on surface, or air), via risk perception
- How mediated by space & time? (e.g. transmission range of pathogen, seasonal contact dependence?)



UML Sequence Diagram

Present for care

Contact enquiry

Contact information

Contact information 2

Contact information 3

Contact information *n*

Collectives

- Groupings are a common multi-scale feature – Herd, Family, Class, Office
- More than the sum of the parts:
 - Can have significant impact on agent perception or behavior
 - Agent may relocate to join new collective
- Common possibilities
 - Purely emergent phenomenon (e.g. herds in CWD example model): Not reified as agent
 - Sometimes epiphenomenal no influence, but instead something that can be used for understanding & analysis
 - Sometimes has very material impact on system behavior
 - Reification as agent (e.g. hierarchical SIR model, gang)
 - Collective can then have own processes & state (e.g. history)

A Multi-Level (Dynamic) Model



Observer Processes

- With an agent-based model, it is often key to have access to many "views" of the model in operation
 - These can aid in validation (calibration, confidence building) and verification (testing), interpretation, results
- The data collected by such observers is typically epiphenomenal – it does not influence the model
- Often there is a significant amount of mechanism & computational effort involved in realizing these
- Detail complexity: significant investment is often further made in visualization interfaces

ODD: 3 Broad Components

- Overview: model goals & high level scope & design
- Design concepts: Different aspects of design being considered
- Details (Remaining elements)

ODD: Remaining Elements

- Initialization
 - Where does initial state come from? Are seeking to make independent of initial state? To test significance of initial state?
- Input data
 - Time series used for model (I think best put in entitiv specification)
- Submodels: Useful abstractions
 - Helpful todescribe early on with broad abstractions (e.g. "partner change", "go to drink", "find food", "stay near mother"
 - Full specification of these are delegated to submodels
 - Seeking low coupling, high cohesion

Sources for Parameter Estimates

- Surveillance data
- Controlled trials
- Outbreak data
- Clinical reports data
- Intervention outcomes studies
- Calibration to historic data
- Expert judgement
- Systematic reviews

Parameter*	Description	Baseline value	Reference
		(units)	
μ	Entry/exit of sexual activity	0.0056 (years ⁻¹)	Garnett and
			Bowden, 2000
с	Partner change rate per	16.08 (years ⁻¹)	Approximated
	Susceptible		from Garnett
			and Bowden,
			2000
β	Probability of infection per	0.70	Garnett and
	sexual contact		Bowden, 2000
φ	Fraction of Infectives who	0.20	Garnett and
	are symptomatic		Bowden, 2000
1/y	Latent period	0.038 (years)	Brunham et.
			al., 2005
$1/\sigma$	Duration of infection	0.25 (years)	Brunham et.
			al., 2005
θ	Asymptomatic recovery	1.5	Garnett and
	coefficient		Bowden, 2000
1/π	Duration of naturally-	1 (year)	Approximated
	acquired immunity		from Brunham
			et. al., 2005



These parameters must have constants specified

ABM Modeling Process Overview



Calibration

 Often we don't have reliable information on some parameters

– Some parameters may not even be observable!

- Some parameters may implicitly capture a large set of factors not explicitly represented in model
- Often we will calibrate less well known parameters to match observed data
 - "Analytic triangulation": Often try to match against many time series or pieces of data at once
- Sometimes we learn from this that our model structure just can't produce the patterns!

Calibration: "Triangulating" from Diverse Data Sources

- Calibration involves "tuning" values of less well known parameters to best match observed data
 - Often try to match against *many* time series or pieces of data at once
 - Idea is trying to get the software to answer the question:
 "What must these (less known) parameters be in order to explain all these different sources of data I see"
- Observed data can correspond to complex combination of model variables, and exhibit "emergence"
- Frequently we learn from this that our model structure just can't produce the patterns!

Calibration

- Calibration helps us find a reasonable (specifics for) "dynamic hypothesis" that explains the observed data
 - Not necessarily the truth, but probably a reasonably good guess at the least, a consistent guess
- Calibration helps us leverage the large amounts of diffuse information we may have at our disposal, but which cannot be used to directly parameterize the model
- Calibration helps us falsify models

Single Model Matches Many Data Sources



Historical Total T2DM Deaths for Time by Ethnicity[GP] : Population Epi Calibrat Total Diabetic Deaths by Ethnicity[GP] : Population Epi Calibration v3 3 T2DM R Historical Total T2DM Deaths for Time by Ethnicity[E]] : Population Epi Calibrati Total Diabetic Deaths by Ethnicity[R]] : Population Epi Calibration v3 3 T2DM Ri

The Pieces of the Elephant Example Model of Underlying Process&Time Series it Must Match



Required Information for Calibration

- Specification of what to match (and how much to care about each attempted match)
 - Involves an "error function" ("penalty function", "energy function") that specifies "how far off we are" for a given run (how good the fit is)
 - Alternative: specify "payoff function" ("objective function")
- A statement of what parameters to vary, and over what range to vary them (the "parameter space")
- Characteristics of desired tuning algorithm
 - Single starting point of search?

Envisioning "Parameter Space" For each point in this space, there

For each point in this space, there will be a certain "goodness of fit" of the model to the collective data



Stochastics in Agent-Based Models

- Recall that ABMs typically exhibit significant stochastics
 - Event timing within & outside of agents
 - Inter-agent interactions
- Can have a pronounced impact on system evolution
- Such stochastics can account for observed patterns that are otherwise hard to explain
- When calibrating an ABM, we wish to avoid attributing a good match to a particular set of parameter values simply due to chance
- To reliably assess fit of a given set of parameters, we need to repeatedly run model realizations
 We can take the mean fit of these realizations
- Often best to match not time series, but summaries

Examples of Stochastics (Compared to Mean Field Deterministic Model)





After 5 replications

After 10 replications

After 40 replications Terminates

ABM Modeling Process Overview



Units & Dimensions

- Distance
 - Dimension: Length
 - Units: Meters/Fathoms/Li/Parsecs
- Frequency (Growth Rate, etc.)
 - Dimension:1/Time
 - Units: 1/Year, 1/sec, etc.
- Fractions
 - Dimension: "Dimensionless" ("Unit", 1)
 - Units: 1

Dimensional Analysis

- DA exploits structure of dimensional quantities to facilitate insight into the external world
- Uses
 - Cross-checking dimensional homogeneity of model
 - Deducing form of conjectured relationship (including showing independence of particular factors)
 - Sanity check on validation of closed-form model analysis
 - Checks on simulation results
 - Derivation of scaling laws
 - * Construction of scale models
 - Reducing dimensionality of model calibration, parameter estimation

Uncertainty Analyses

- Same relative or absolute uncertainty in different parameters may have hugely different effect on outcomes or decisions
- Help identify parameters that strongly affect
 - Key model results
 - Choice between policies
- We place more emphasis in parameter estimation into parameters exhibiting high sensitivity

Uncertainty Analysis: Initial Value

- Frequently we don't know the exact state of the system at a certain point in time
- A very useful type of sensitivity analysis is to vary the initial value of model stocks
- In Vensim, this can be accomplished by
 - Indicating a parameter name within the "initial value" area for a stock
 - Varying the parameter value

Robustness Analysis

- To what degree are model conclusions robust under changing model structural and other large assumptions?
 - Distinguish cases where
 - Results depends on something essential about the model
 - Results depend on happenstance of simplifying assumptions
 - e.g. spatial neighborhood assumption, size or granularity of space, convenient assumptions regarding rules or what is known
- We want to rule out cases where getting "right result for wrong reasons"!
- Seek to find whether conclusions change radically when just a few assumptions are changed?
- Process is similar to what used for submodel testing, but done for entire model

Imposing a Probability Distribution Monte Carlo Analysis

- We feed in probability distributions to reflect our uncertainty about one or more parameters
- The model is run many, many times (realizations)
 - For each realization, the model uses a different draw from those probability distribution
- What emerges is resulting probability distribution for model outputs

Example Resulting Distribution



Static Uncertainty



Dynamic Uncertainty: Stochastic Processes



This is a form of sensitivity analysis, but because we are capturing effects of model stochastics – rather than our lack of knowledge, we don't term "uncertainty analysis"

Dynamic Uncertainty: Stochastic Processes



Mathematical Analysis of Models

System Linearization (Jacobian)



Example: Simple SITS Model



Associated System of State Equations



ABM Modeling Process Overview



Contact Tracing Simulation

Run the model and switch to Main view

We can make it better!

Network type	Network Settings	Parameter Settings
Random	Connect Per Agent	Simulation Fraction of RI
C Small world	Notes: Connects Per Agent is for Random and Small World Net	Norks Simulation Fraction of NonRI
C Scale free	Neighbourhood Link Prob Notes: Link Prob is for Small World Networks	Enable Database
	ScaleFreeM	
	Note: ScaleFreeM is for Scale Free Networks	
Contact Tracing Policy Se	lection	
No Contact Tracing Pro	ogram 🐴	
C Contact Tracing With F	Priority	
Contact Tracing Priority S	iettings (Weight)	
🗹 Age Priority 🔽 Ethnic	ity Priority 🔽 RR of Count Priority	
Contact Tracing Targets		
Tracing Infectious Acti	ve TB Cases ONLY	
O Tracing All Active TB Cases		
C Tracing Infectious Acti	ve TB Cases and Primary TB	
Contact Tracing Percenta	ge on Average	
Average Percentage of Scenario Information	f Contacts to Investigate:	
Description		

Scenario Results (Means)



SO (baseline=No Contact Tracing)

- S1(Target=Infectious&PrimaryTB_Lost=30to40%_NoPriority_TracingFraction=90%)
- S2(Target=Infectious&PrimaryTB_Lost=10%_NoPriority_TracingFraction=90%)

Variability in Results

	Cumulative Incident Cases (Active TB)				
Scenario Id	Mean	Max	Min	Std. Deviation	C.V
S_0	425.633	614	289	74.659	0.175
S_1	311.767	429	217	49.646	0.159
S_2	279.1	392	211	49.682	0.178
S_3	318.667	403	207	48.093	0.151
S_4	283	364	193	40.403	0.142
S_5	302.233	486	194	64.917	0.215
S_6	363.2	508	239	70.19	0.193
S_7	291	383	190	53.018	0.182
S_8	265.5	400	185	44	0.166
S_9	315	438	184	49.2	0.156
S_{10}	271.6	387	192	41.57	0.153

ABM Modeling Process Overview



Contact Tracing Simulation

Run the model and switch to Main view

We can make it better!

Network type	Network Settings	Parameter Settings
Random Small world	Connect Per Agent Notes: Connects Per Agent is for Random and Small World No	Simulation Fraction of RI
C Scale free	Neighbourhood Link Prob Notes: Link Prob is for Small World Networks ScaleFreeM	
Contact Tracing Policy Se	Note: ScaleFreeM is for Scale Free Networks election	1.1 1 + 1. 1.1
• No Contact Tracing Pr	rogram 🔰	TIA T. C.A
C Contact Tracing With Priority		
Contact Tracing Priority §	Settings (Weight)	
🗹 Age Priority 🔽 Ethnia	city Priority 🔽 RR of Count Priority	
Contact Tracing Targets		
⊙ Tracing Infectious Act	ive TB Cases ONLY	
C Tracing All Active TB Cases		
C Tracing Infectious Active TB Cases and Primary TB		
Contact Tracing Percenta	age on Average	
Average Percentage o Scenario Information	of Contacts to Investigate:	Z
Description		

Key Take-Home Messages from this Lecture

- Models express dynamic hypotheses about processes underlying observed behavior
- Frequently observed behavior is "emergent" it is qualitatively different than the behavior of any one piece of the system, or a simple combination of behavior of those pieces
- Models help understanding how diverse pieces of system work together
- ABM focus on agent interactions as the fundamental shapers of dynamics
- Models are specific to purpose