



METIER Graduate Training Course
“Ecological Modelling”
22 May – 2 June 2008, Leipzig & Bad Schandau (Germany)

Lecture 6: ***Bridging the Gap between Bottom-up & Top-down Modeling Approaches***

Part I

Justin M. Calabrese
Helmholtz Centre for Environmental Research – UFZ, Leipzig
Department of Ecological Modeling



Justin Calabrese; Lecture 6 „Micro to macro connections“ (Part I; Day 6)

A question of perspective and scale

What is the appropriate level of biological organization to focus on when building models?



Photo by OSF/A. Shay/Animals Animals—Earth Scenes

Individuals?
Bottom up?

Populations?
Top down?

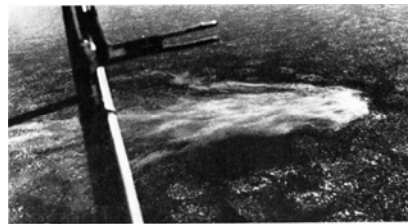


Photo by H.J. Sayer

A question of perspective and scale

Individual Level:



Any given individual, when viewed in isolation, may appear to behave erratically.

Population Level:



A collection of such individuals may display statistical regularities → patterns.

A question of perspective and scale

The answer, of course, depends on the research question, study goals, etc...

Classical equation-based modeling and individual-based simulation modeling largely differ in how they answer this question

Both approaches have their strengths and weaknesses

Though these approaches are often presented as being diametrically opposed, connections between them sometimes exist

What can we learn from such cases?

Individual-based (oriented) approach

Loose definition

The essence of the individual-based approach is the derivation of the properties of ecological systems from the properties of the individuals constituting these systems

-Adam Lomnicki 1992

Individuals obey (often stochastic) behavioral rules

Usually implemented as a computer simulation model

“Bottom-up” strategy to describing ecological phenomena

Strengths of bottom-up approach

Focuses on Individuals:

- Strong arguments in favor of doing so
- System properties "emerge" from behavior of inds.

Flexible:

- Many of the constraints imposed by classical approaches no longer relevant
- Fewer artificial assumptions required

Lower Barrier to Entry:

- Easier to get started with
- Closer to the way empirical ecologists think about problems

Weaknesses of bottom-up approach

Communication:

- Often difficult to clearly explain model structure
- Only complete description is the source code

Lack of Standards:

- Solutions often ad-hoc
- Little reuse of standardized code

Analysis and Statistics:

- Model analysis can be difficult for complicated models
- Often difficult to connect IBMs to data in a formal way (i.e., can't write down a likelihood for the model)

The middle ground

The strengths of the top down approach are often the weaknesses of the bottom up approach and vice versa. *These approaches are often complementary.*

Suggests that finding a direct link between them could be very productive.

Micro-scale model gives macro-scale model context and provides a direct link to individual behavior.

Macro-scale version facilitates analysis and communication of results.

How do we find a connection?

What is the relationship between parameters of micro-scale model and those of macro model?

Justin Calabrese; Lecture 6 „Micro to macro connections“ (Part I; Day 6)

8

Micro-scale to macro-scale connections

The essence of the individual-based approach is the derivation of the properties of ecological systems from the properties of the individuals constituting these systems

-Adam Lomnicki 1992

Micro to macro connections:

In certain situations it is possible to establish this connection mathematically (often by approximation)

Use statistical regularities that "emerge" from behavior of many inds. to avoid having to keep track of every ind.

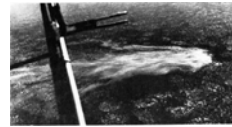
Model the dynamics of these summary quantities

A key advantage of this approach is that it establishes a clear relationship between the parameters of the micro and macro-scale models. In many cases, they are exactly the same.

Justin Calabrese; Lecture 6 „Micro to macro connections“ (Part I; Day 6)

9

The rift to be bridged



These approaches tend deal in different currencies so one of the main challenges is to find “exchange rates” between them

Discrete entities

?

Aggregate state variables

Discrete time

Continuous time

If/Then rules



Instantaneous rates

Stochastic

Deterministic

Examples

Though it is not possible to establish such connections for arbitrarily complicated models, connections exist for certain classes of widely used ecological models:

Grid-based (lattice) population models

Continuous space population models

Random walks

Stochastic patch occupancy metapopulation models

Examples

Notice that all of these examples are "spatial" models in some sense

This is good as one of the biggest limitations of classical equation-based modeling is that spatial structure is often ignored

Means that we can often use this approach to study spatial ecological questions:

What is the effect of spatial structure on the systems' dynamics?

How and when does spatial structure form?

What kinds of spatial structures are possible and how are these related to model parameters?

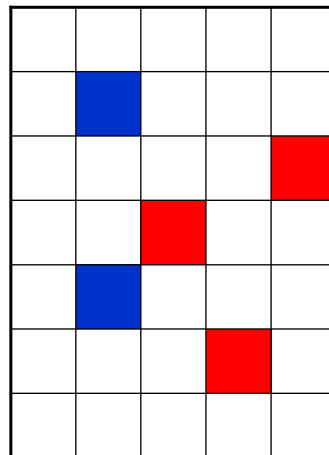
Stochastic cellular automata

Simulation universe is a grid of cells

Cells are characterized by discrete states

Probabilistic (often context dependent) rules govern state transition

Not truly "individual-based" but flexible and very widely used (pred/prey example)



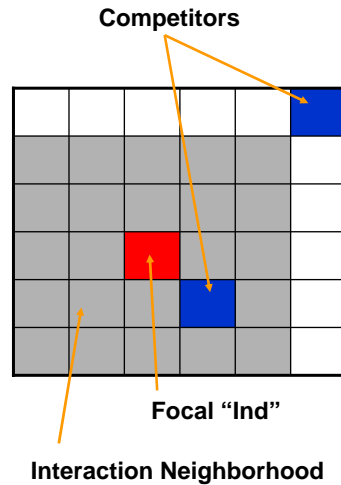
Stochastic cellular automata

Often used as basis for spatially explicit models

Local interactions occur over defined "interaction neighborhoods"

Can include multiple species, complex interactions, movement, non-homogeneous habitats, etc...

Can accommodate more complicated lattice structure



A simple spatially explicit pop. model

Simple single species birth/death process with spatially limited dispersal and local (neighborhood) competition.

A great many grid based models in ecology are some variation on this basic theme.

Goals:

Show logic and basic techniques involved in establishing a micro-macro connection

Demonstrate relationship between this model and a well-known classical model

Illustrate how classical models can (sometimes) be better justified by deriving them in this way

Discrete individuals or densities?

In the CA simulation model, we keep track of discrete individuals (cells)

In differential equations, state variables are continuous (e.g., density)

Solution: Use proportion of occupied cells as main state variable of system

$$\rho_1 = \frac{N}{S}$$

where N is number of individuals and S is number of lattice sites.

Time: Continuous or discrete?

One of the first obstacles you'll encounter is how to model time

Discrete time is more natural for simulation modeling (timesteps)

Continuous time is more natural for equation-based modeling (calculus)

Simulations often use discrete timesteps without any real justification for it (hrs., days, weeks, months, years, etc...)

Switching to (approximating) continuous time can be computationally expensive (must simulate very small time increments), but can greatly increase mathematical tractability. Efficiency can be improved by using asynchronous updating

Transitions: Rates or probabilities?

Given that we've decided on cont. time and differential equations, how do we deal with transitions?

In stochastic cellular automata, transitions happen probabilistically according to transition rules:

In diff. eqn. models, transitions happen continuously (and deterministically) and are characterized by instantaneous rates: death rate = α

What is the connection?

```
If random() < probDie
    die
End If
```



$$probDie = 1 - e^{-\alpha dt}$$

Ind. stochasticity vs. pop.-level determinism

Individuals, though characterized by the same behavioral rules, can display high variability...

...but we (usually) need a deterministic description of the population

Solution:

The average of many individuals often behaves deterministically

Track the averages of state variables (i.e., ρ_1)

Track the average conditions (i.e., neighborhood density) in which an individual finds itself

Simulation model description

Cells can be in either state 1 (occupied) or 0 (unoccupied).

Habitat is homogeneous (implies no spat. variation in trans. rules)

Time proceeds in small steps of length dt .

Periodic (toroidal) boundary conditions

Interaction neighborhood is same for birth and death and
 Its size is $z = 8$

Simulation model description

Birth:

An occupied cell sends out offspring over its interaction neighborhood at rate b ($p_b = 1 - e^{-bdt}$).

If an offspring lands on an already occupied cell nothing happens, if it lands on an unoccupied cell, it establishes.

Death:

An occupied cell dies at some intrinsic rate α if its interaction neighborhood is completely unoccupied.

Each neighbor increases the focal cell's death rate by a factor δ ($p_d = 1 - e^{-(\alpha + \delta C)dt}$).

Summary

What we have:

A simple stochastic CA we can simulate

"Exchange rates" between the different currencies used in CA and differential eqns.

We have already made the biggest conceptual step

What's left:

Basically, a bunch of bookkeeping

How do we proceed:

Focus on frequencies of different cell state configurations

Consider how these summary stats change when ind cells change

Matsuda et al. 1992, Dieckmann et al. 2000, Ellner 2001

Three types of interrelated frequencies

Singlet frequencies:

$$\rho_1 = \frac{N}{S} \qquad \rho_0 = 1 - \rho_1$$

Pair frequencies:

$$\rho_{ij} = \text{prob. that a neighbor pair is in state } ij$$

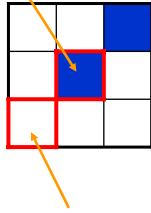
Conditional pair frequencies:

$$q_{j/i} = \text{prob that neighbor is in state } j \text{ given focal site in } i$$

Pair frequencies

ρ_{ij} = prob. that a neighbor pair is in state ij

Pick a cell on the lattice at random



Randomly pick one of its neighbors

The probability that the first site is in state i and the second is in state j is ρ_{ij}

Pair freqs. obey rules of probability

Four possible pair frequencies, they must sum to one, so only three are independent

$$\rho_{11} + \rho_{10} + \rho_{01} + \rho_{00} = 1$$

Because of neighborhood symmetry,

$$\rho_{10} = \rho_{01}$$

This means simply that if I am your neighbor, you must be my neighbor.

Given these constraints, only two are actually independent.

Pair freqs. obey rules of probability

Now consider relationships between singlet and pair frequencies:

$$\rho_1 = \rho_{11} + \frac{1}{2}(\rho_{10} + \rho_{01})$$

But because of neighborhood symmetry

$$\rho_1 = \rho_{11} + \frac{1}{2}(2\rho_{10})$$

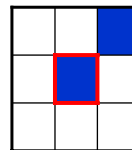
So we have

$$\rho_1 = \rho_{11} + \rho_{10}$$

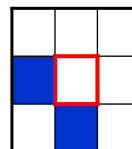
Local densities

The conditional probabilities $q_{1/1}$ & $q_{1/0}$ are sometimes called "local densities" because they represent the average neighborhood conditions around occupied and unoccupied cells, respectively.

$$q_{1/1} = \frac{1}{8}$$



$$q_{1/0} = \frac{1}{4}$$



Local densities obey rules of probability

Local density terms are conditional probabilities. Applying the definition of conditional probability we see that

$$q_{j/i} = \frac{\rho_{ij}}{\rho_i} = \frac{\rho_{ji}}{\rho_j}$$

With that, we have all we need to describe how state transitions in the simulation model happen in the currencies used in diff. eqn. models

The protocol

We want an expression that looks like

$$\frac{d\rho_1}{dt} = \text{total birth rate} - \text{total death rate}$$

The recipe to get it:

- 1) Consider a particular state transition (e.g. $1 \rightarrow 0$)
- 2) Find the appropriate transition rate for a single site
- 3) Find the average transition rate across all sites
- 4) Obtain the total transition rate by multiplying. by the appropriate singlet or pair freq.
- 5) Substitute total transition rates into above eqn.

Death

Consider the $1 \rightarrow 0$ transition (death) for a single site

$$\text{death rate} = \alpha + \delta C$$

Where α is intrinsic death rate, δ is the contribution of each neighbor to the death rate and C is number of neighbors.

The average death rate of a site in state 1 depends on the average number of neighbors, or

$$\text{avg. death rate} = \alpha + \delta \bar{C}$$

Death

What is \bar{C} in terms of the rules of the model?

$$\bar{C} = z q_{1/1} \quad \text{so then}$$

$$\text{avg. death rate} = \alpha + \delta z q_{1/1}$$

$$\text{Define} \quad d = \delta z$$

$$\text{avg. death rate} = \alpha + d q_{1/1}$$

Final step is to mult. above by avg. freq. of sites in state 1

$$\text{total death rate} = (\alpha + d q_{1/1}) \rho_1$$

Birth

So where are we now?

$$\frac{d\rho_1}{dt} = \text{total birth rate} - (\alpha + d q_{1/1})\rho_1$$

Next consider $0 \rightarrow 1$ transition (birth) for a single site

$$\text{ind. birth rate} = \frac{b}{z} C$$

Again, average rate depends on avg. number of neighbors

$$\text{avg. birth rate} = \frac{b}{z} \bar{C} = \frac{b}{z} z q_{1/0} = b q_{1/0}$$

Birth

Multiply by prop. of unoccupied site to obtain total rate

$$\text{total birth rate} = b q_{1/0} (1 - \rho_1)$$

Substitute this into our diff. eqn.

$$\frac{d\rho_1}{dt} = b q_{1/0} (1 - \rho_1) - (\alpha + d q_{1/1})\rho_1$$

Hooray!!!

So we're done, right?

Unfortunately, no. But why not?

$$\frac{d\rho_1}{dt} = b q_{1/0} (1 - \rho_1) - (\alpha + d q_{1/1}) \rho_1$$

Some more bookkeeping...

From the death term, the pair prob.

$$q_{1/1} \rho_1 = \frac{\rho_{11}}{\rho_1} \rho_1 = \rho_{11}$$

is unknown.

So we're done, right?

Similarly, in the birth term, the pair prob.

$$q_{1/0} (1 - \rho_1) = \frac{\rho_{10}}{\rho_0} (1 - \rho_1) = \frac{\rho_{10}}{(1 - \rho_1)} (1 - \rho_1) = \rho_{10} = \rho_1 - \rho_{11}$$

To go further we need another equation for the pair corr. that looks like

$$\frac{d\rho_{11}}{dt} = \text{total pair birth rate} - \text{total pair death rate}$$

No problem, through similar bookkeeping, we'll figure it out...

Moment equations

An open hierarchy of equations of the form

$$\begin{aligned}\frac{d\rho_1}{dt} &= f(\rho_1, \rho_{11}) \\ \frac{d\rho_{11}}{dt} &= f(\rho_1, \rho_{11}, \rho_{111}) \\ \frac{d\rho_{111}}{dt} &= f(\rho_1, \rho_{11}, \rho_{111}, \rho_{1111}) \\ &\vdots\end{aligned}$$

By itself, this is useless. To get anywhere we have to break this chain of dependencies and "close" the hierarchy.

In other words we need to find an approximate moment closure.

The simplest possible closure

The easiest way to deal with the closure problem is to simply throw away all the spatial information in the model, this is the "mean-field" approximation:

$$q_{1/1} \text{ or } q_{1/0} \rightarrow \rho_1$$

Going back to our model

$$\begin{aligned}\frac{d\rho_1}{dt} &= b q_{1/0} (1 - \rho_1) - (1 + d q_{1/1}) \rho_1 \\ &\quad \downarrow \qquad \qquad \qquad \downarrow \\ \frac{d\rho_1}{dt} &= b \rho_1 (1 - \rho_1) - (1 + d \rho_1) \rho_1\end{aligned}$$

Truncating the hierarchy

What we have effectively done is to break the hierarchy after the first equation, thus we retain only information about the mean density

$$\frac{d\rho_1}{dt} = f(\rho_1, \rho_{11}) \quad \text{Mean Density}$$

$$\frac{d\rho_{11}}{dt} = f(\rho_1, \rho_{11}, \rho_{111}) \quad \text{Spatial Structure}$$

$$\frac{d\rho_{111}}{dt} = f(\rho_1, \rho_{11}, \rho_{111}, \rho_{1111})$$

⋮

Unfortunately, we lose a lot of information by doing so

The logistic equation

After some algebra we have

$$\frac{d\rho_1}{dt} = (b - \alpha)\rho_1 - (b + d)\rho_1^2$$

which is recognizable as the logistic equation. Setting it equal to zero and solving for ρ_1

$$\rho_1^* = \frac{b - \alpha}{b + d}$$

gives the equilibrium proportion of occupied sites as a function of the parameters.

Recovering the classical logistic

Start with

$$\frac{d\rho_1}{dt} = (b - \alpha)\rho_1 - (b + d)\rho_1^2$$

define $r = b - \alpha$ and recall $\rho_1 = \frac{N}{S}$

then

$$\frac{dN}{Sdt} = r \frac{N}{S} - (b + d) \left(\frac{N}{S} \right)^2$$

and multiplying through by S

$$\frac{dN}{dt} = rN - (b + d) \frac{N^2}{S}$$

Recovering the classical logistic

Recall $\rho_1^* = \frac{b - \alpha}{b + d} = \frac{K}{S}$

Notice that K , the carrying capacity can be defined in terms of the parameters and the grid size. In other words, K is a consequence of the model's structure and its parameter values, not an arbitrary number. Now solve for S

$$S = \frac{(b + d)K}{r} \quad \text{and subs. this into above} \quad \frac{dN}{dt} = rN - \frac{r(b + d)N^2}{(b + d)K}$$

and finally

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right)$$