Mathematical Techniques in Evolution and Ecology

How to construct a model

Based on chapter 2 in Otto and Day (2007)

Spring Quarter 2015
Simon Aeschbacher
saeschbacher@ucdavis.edu

Outline

Goals

- To describe the steps involved in developing a model
- To derive equations that describe the dynamics of a biological phenomenon

Concepts

- Discrete-time model
- Continuous-time model
- Recursion and difference equations
- Differential equations
- Life-cycle diagrams
- Flow diagrams
- Mass action

Types of models

In this course, we focus on *dynamical* models, which describe how a system changes *over time*. From a technical point of view, these can be categorised according to the following criteria:

- *Deterministic* versus *stochastic* models
- *Discrete*- versus *continuous*-time models
- *Dimensionality* (the number of dynamical variables / equations)

Obviously, these criteria are not mutually exclusive and the list is not complete.

Ingredients and parsimony

Ingredients to a dynamical model

- **Variables**: quantities that *change over time*
- **Dynamics**: *pattern of change* over time
- **Parameters**: quantities that remain *constant over time*
- **Dimensionality** (the number of dynamical variables / equations)

The principle of parsimony

- *Prefer a simple explanation* or model over a complex one, if both are *equally compatible with the data*
- Start from a simple null model; rejected in favour of a more complicated model if necessary

Seven steps to modelling a biological problem

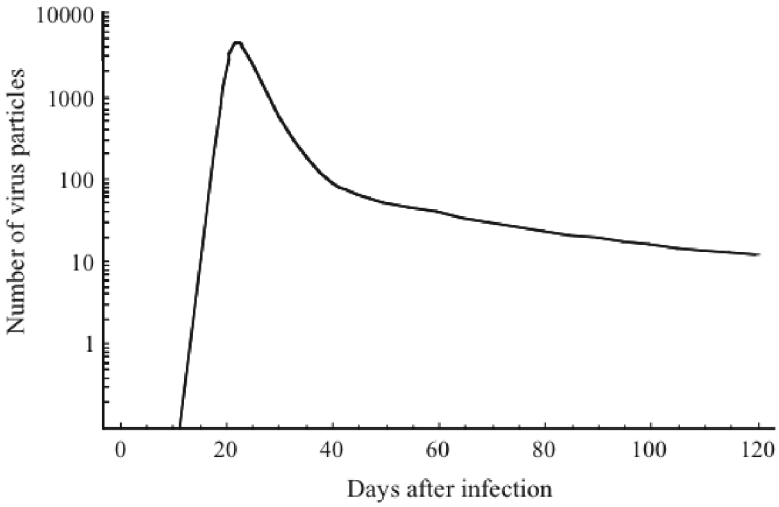
In Box 2.1, Otto and Day (2007) [OD07] propose seven steps to modelling a biological problem:

1. Formulate the question
2. Determine the basic ingredients
3. Qualitatively describe the biological system
4. Quantitatively describe the biological system
5. Analyse the equations
6. Checks and balances
7. Relate the results back to the question

In the following, we will discuss these steps in more detail.

1. Formulating the question

- What do you want to know? Describe the model in the form of a question.
- Start from an interesting observation. There may be an unexpected resolution!
  - Example: Why does the # of HIV particles drop after an initial peak?



- Start with the simplest, biologically reasonable description of the problem
  - Example: “The immune system responds” (complex) vs “HIV runs out of CD4+ cells to infect” (simple) [See OD2007 for details]

Three examples that will be revisited

- How does the number of branches of a tree change over time?
  → *Population growth*
- How does cycling affect the number of squirrels on campus?
  → *Immigration*
- How does the number of people with the flu change over the season?
  → *Interactions*

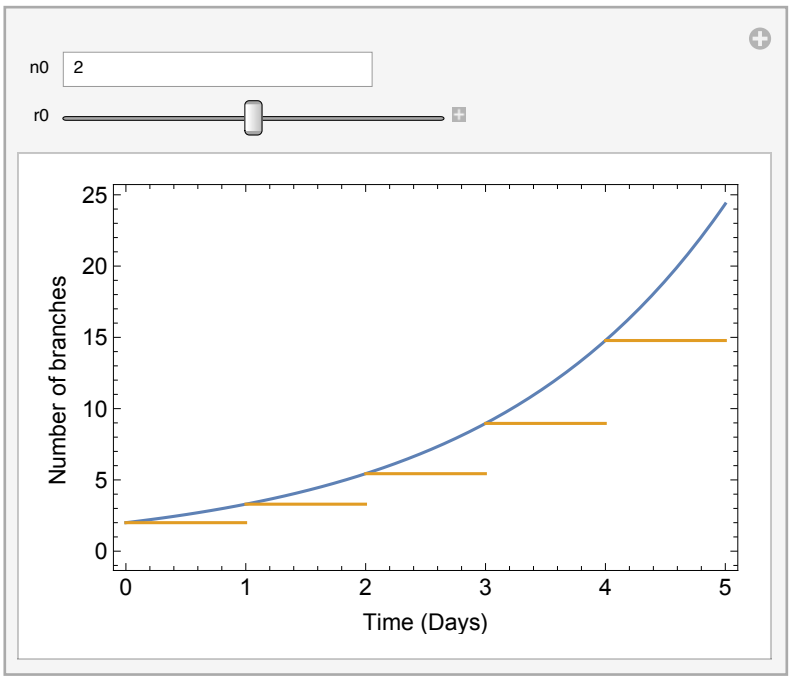
2. Determining the basic ingredients

- Define the variables. Start simple, add more if necessary. Use letters / short expressions.
  - Branches (squirrels): $n(t)$ for the # of branches (squirrels on campus) at time t
  - Flu: $n(t)$ for the # of people with the flu, $s(t)$ for the # of susceptible people
  - In practice, it is often helpful to treat variables as continuous
    - Branches (squirrels): interpret the # of branches (squirrels) as 'biomass'
    - Flu: the # of virus particles is likely large; treating them as continuous introduces little error
- Describe constraints on the variables
  - Branches, squirrels: $n(t) \geq 0$
  - Flu: $n(t), s(t) \geq 0 \ \bigwedge \ n(t) + s(t) = N$, where N is the total population size
- Describe interactions among variables

- Discrete or continuous time?
  - **Discrete-time models** track changes to variables in discrete *time steps*
  - **Continuous-time models** allow variables to *change at any point in time*
  - Choose a *time scale* (discrete time: time step; continuous time: unit of rates)
    - Examples: discrete time step of *1 day*; rates measured in units *per day*

In[8]:= **mpContDiscDyn**

Out[8]=



- It is sometimes *easier to think in terms of discrete time*, but *continuous-time* models are often *easier to analyse mathematically*. Fortunately, it is often possible to derive a continuous-time version of a discrete-time model.
- Discrete time:
  - **Recursion equations**, describing the *value* of a variable *in the next time step*

$$n(t+1) = \text{"some function of } n(t)\text{"} \tag{1}$$

  - **Difference equations**, describing the *change* in a variable *per unit of time*

$$\triangle n(t) = n(t+1) - n(t) = \text{"some function of } n(t)\text{"} \tag{2}$$

- Continuous time:
  - **Differential equations**, describing the *rate of change over time* of a variable

$$\frac{dn(t)}{dt} = \text{"some function of } n(t)\text{"} \tag{3}$$

- Define the parameters
  - Remain *fixed over time* as the variables change
  - Common to use italicised roman letters or lower-case greek letters
  - Parameters *representing events per unit time*:
    - Discrete-time models: *number of events* per time step
    - Continuous-time models: *rate of events* per unit time
  - Other parameters (e.g. the probability of a certain type of event) retain the same definition in discrete and continuous time

- Examples:

In[9]:=
```
paramExampleTable
```

Out[9]=

Discrete time	Continuous time
b, nr of branches budding off each old branch per day	*β*, *rate* of budding for each old branch per day
d, fraction of squirrels killed by cyclists per day *b, number* of squirrels born per squirrel per day	*δ*, *rate* of death due to cyclists per day *β*, *rate* of birth per squirrel per day
c, fraction of healthy people exposed to flu carriers per day *a, probability* of transmission upon exposure	*γ*, *rate* of contact between carriers and susceptibles per day *a*, *probability* of transmission upon exposure

- Describe constraints on the parameters
  - Can a parameter be negative?
  - Does it describe a fraction, proportion, or probability?
    - Example of squirrels: Parameter d is a *fraction*, i.e. $0 \leq d \leq 1$ must hold. However, δ is a *rate* and can have any positive value. Note the difference between the discrete- and continuous-time model here!
  - Prior knowlegde? What is biologically reasonable? Assumptions?
    - Example: The # of new branches growing from an existing branch per day is likely to be very small, i.e. $b \ll 1$.
  - **Remark:** Constraints in the form of assumptions about the value of a parameter are often crucial to obtaining mathematical results. It is therefore very important to keep track of these assumptions and be aware of whether or not the results apply to a given range of parameter values.

3. Qualitative description of the system

- **Life-cycle diagrams** illustrate the *order of events within* each time step (for discrete-time models only!)
  - Draw a cycle (i.e. circle) for each variable
  - On each cycle, mark all events that can happen within a given stime step (e.g. a day), including a ‘census’ event
  - Order the events such that they are biologically meaningful (it may matter!) and consistent across cycles
  - Use primes to distinguish variables before and after a given event

- **Flow diagrams** illustrate how each variable *affects* its own *dynamics* and those of other variables
  - Draw a circle for each variable
  - Draw a *returning* arrow if the variable influences *itself*
  - Draw arrows *entering* the circle for effects that *increase* the value of the variable (influx)
  - Draw arrows *exiting* the circle for effects that *decrease* the value of the variable (eflux)
  - Use *merging* arrows for *interactions* between variables
  - Label arrows by the flow they represent, including how the flow depends on the variable(s)
  - Pay attention with discrete-time models including multiple events per time step: keep track of and stick to the order of events
  - Interactions can *sometimes* be modelled according to the **mass-action principle**, which assumes that the rate of interaction between two variables is proportional to the values of each

- **Tables of events** are useful for discrete-time models with multiple interactions
  - Include one row for each type of interaction
  - Include one columns for the interaction, one for the number of contacts, and one for the outcome of the interaction w.r.t. to each variable

eventsTable

Interaction	# of contacts	Result for infected	Result for susceptible
Infected × infected	$c\, n(t)\, n(t)$	No change	No change
Infected × susceptible	$c\, n(t)\, s(t)$	$+a$	$-a$
Susceptible × susceptible	$c\, s(t)\, s(t)$	No change	No change

>> Let us do it for some of our examples...

4. Quantiative description of the system

- Using the diagrams/tables from step 3, write down the equations
  - Discrete time
    - Recursion equations:

$$n(t+1) = n(t) + \text{sum of increases} - \text{sum of decreases} \tag{4}$$

    - Difference equations:

$$\triangle n = \text{sum of increases} - \text{sum of decreases} \tag{5}$$

    - **Important:** Specify order of events and stick to it
  - Continuous time
    - Differential equations:

$$\frac{d(n(t))}{dt} = \text{total rate of increase} - \text{total rate of decrease} \tag{6}$$

    - No need to worry about order of events

>> Let us do it for some of our examples...

- Perform checks:
  - Are constraints on variables still met?
  - Do the units on the right- and left-hand side match?
- Will results from the model address the question?

Recipe 2.1: Recursion equations from life-cycle diagrams (discrete time)

For each variable n:

1. Use $n'(t)$, $n''(t)$, $n'''(t)$, etc. to denote the value of the variable after the first, second, third, etc., event in the life cycle, and obtain recursions for these according to Eq. (4).
2. Set $n(t+1)$ to the value of n after the final event in the life cycle.
3. Substitute the recursion for $n'(t)$ into the recursion for $n''(t)$ and simplify. Then substitute the recursion for $n''(t)$ into the recursion for $n'''(t)$ and simplify, etc., until the resulting expression gives a recursion for $n(t+1)$ solely in terms of $n(t)$.

Recipe 2.2: Difference equations from recursion equations (discrete time)

For each variable n:

1. Calculate $n(t+1)$ using Recipe 2.1.
2. Subtract $n(t)$ from $n(t+1)$ and simplify to get the difference equation, $\Delta n = n(t+1) - n(t)$, describing the change in n per time step.

Recipe 2.3: Recursion equations from flow diagrams (continuous time)

For each variable n:

1. Considering each arrow in turn, update the value of each variable by taking its previous value
   - plus the flow if the arrow enters the circle of n
   - plus the flow if the arrow leaves and returns to the circle
   - minus the flow if the arrow leaves the circle
2. Set $n(t+1)$ to the value of n after the final arrow has been considered

Recipe 2.4: Differential equations from flow diagrams (continuous time)

For each variable n:

1. $\frac{d(n(t))}{dt}$ = the flow rates along arrows entering the circle
+ the flow rates along arrows leaving and returning to the circle
– the flow rates along arrows exiting the circle.

5. Analysing the equations

Much of the rest of the course will be devoted to this. Some of the techniques are:

- Graphical and numerical analyses (unit 4; chapter 4 in OD2007)
- Equilibrium and stability analyses (units 5, 7, and maybe 8; chapters 5, 7, and 8 in OD2007)
- Deriving general solutions (units 6, and maybe 9; chapters 6 and 9 in OD2007)
- Determining long-term or asymptotic behaviour (chapter 10 in OD2007)
- Analysing the model for periodic behaviour (chapter 11 in OD2007)

6. Checks and balances

Modelling is often an **iterative process**. Mistakes may cause you to start over. It is good to get into the habit of checking results, using other pieces of information and asking a few questions.

- Check if the results obey the *constraints* on the variables you identified previously
  - Examples: Do proportions sum to 1? Do you obtain a negative value where you don't expect it?
- Make sure each equation has the *right units* (dimensionality) on both sides
  - Do this check for a number of equations we derived for our three examples
- Look at *special cases* where you know what should happen. Often, this involves setting (some) parameter(s) to 0.
  - Examples: In the squirrel model, setting $d = m = 0$ means we consider the special case of no migration and no deaths due to cyclists. The model then collapses to the tree-branch model, and this is simply a model of exponential growth, which is well-known.

- Validation: Do the results *make sense*, i.e. are they *plausible*?
- Is the model *too simple* or *too complex*?
  - Tree-branch example: Infinite growth at the same rate is unrealistic.
  - Flu example: Include a class of 'recovered and resistant' people.
- Can you *group some parameters*?
  - Flu example: $a\,c \to f$, i.e. infectivity = [contact rate] × [probability of infection per contact]
  - In view of inference base on a model, this is related to the question of whether or not parameters will be '*identifiable*'. In the flu example, we would say that "a and c are confounded", or that "a is not identifiable without knowledge about c", and vice versa.

7. Relating equation(s) back to question(s)

- Do the results answer the biological question?
- Are the results counterintuitive?
- What are the predictions? Describe potential experiments to test these.
- Think of observations (data) that can be better explained and understood after the modelling

Problem assignment

1. In the squirrel example, assume that death through cyclists occurs before immigration, and immigration is followed by birth. Derive the discrete-time equation (recursion) and compare it to the one we obtained above. Start from the life-cycle diagram.
2. [Problem 2.5 in OD2007] Consider a model of disease transmission with the following equations: $\frac{dS}{dt} = \theta - d\,S - \beta\,S\,I + \gamma\,I$, and $\frac{dI}{dt} = \beta\,S\,I - (d + \nu + \gamma)\,I$. The variables S and I denote the number of susceptible and infected individuals. (a) Draw and label a flow diagram for these two variables. (b) Suggest a plausible biological interpretation of the parameters γ and ν.
3. [Problem 2.6 in OD2007] In the flu model, suppose that after contracting the flu, people are initially resistant to reinfection, but this immunity eventually wanes. (a) Alter the flow diagram for the flu model that we derived to include a “recovered and immune” class with these properties. (b) Suppose that immune individuals have a constant per capita rate of losing immunity. What are the continuous-time equations for this modified model?