Computational biology, 2023

Teachers Kristian Gustafsson, Bernhard Mehlig, Ludvig Storm (problem sessions)

Canvas course page https://chalmers.instructure.com/courses/22434

Literature and references

Lecture notes will be made available.

Course book 'Mathematical Biology I', by J. D. Murray (at Cremona)

Selected chapters from

- 'Nonlinear Dynamics and Chaos', by Strogatz

- 'Diffusion and Ecological Problems', by Okubo and Levin

- 'Calculating the Secrets of Life', by Lander and Waterman

- 'Mathematical Biology II', by Murray

Examination

Three sets of hand-in problems + exam on campus (you must sign up before Feb. 26). To pass the course a pass grade on both the problem sets and the exam are required (see course page).

Piazza discussion forum

Forum for questions about homework or lectures. Questions should be asked here rather than in emails to the teachers. This allows other students to answer questions and to see answers to the questions.

Contents of the course   This course introduces mathematical modeling of complex biological systems, such as

- Population growth

- Interacting species and reaction kinetics

- Spreading of diseases (epidemiology: influenza, AIDS, etc.)

- Population genetics

- Morphogenesis (pattern formation, animal coating)

- Synchronization

Some of the methods we use to model and analyze the systems are

- Discrete and continuous dynamical systems

- Stochastic approaches

- Reaction-diffusion equations

- Time-series analysis

1 Population dynamics (Murray Ch. 1,2)

1.1 Introduction

The dynamics of biological populations (anything from microbes to humans) is in general a complicated problem influenced by many factors such as

- population size

- reproduction pattern

- geographical distribution

- mutation

- migration

- natural selection

It is very difficult to analyze the combined effect of these (and more) factors on the evolution of a population. In practice it is therefore necessary to set up a model that only includes a small number of essential factors that influence the population. There are several types of models:

- **Experimental models** For example a lab experiment with cultures of bacteria isolated from their surroundings (Petri dish).

- **Computer models** For example simulating the dynamics of a population given a set of numerical rules. One example could be agent-based modelling to simulate interactions between individuals in a population.

- **Mathematical models** Defined by a set of hypotheses that specifies relations between observables and parameters. For example equations of motion in Physics or the growth equations in the following sections.

There are several ways in which Mathematical models may be useful:

- They may reveal which parameters or mechanisms are most significant and may thus suggest critical experiments (experiment that determines whether a hypothesis is superior to all other accepted hypotheses in a scientific community)

- They may serve as guide to collection, organization, and interpretation of data

- They make quantitative predictions of the behaviour of a system that can be confirmed in experiments.

Computer models are often useful in situations where a number of elementary rules appear reasonable, but at the same time the behaviour of the system appears too complicated (or chaotic, or irregular), in other words too complex to cast in terms of a few elementary mathematical relations between observables.

Some obvious requirements for a model to be meaningful are:

- it must be possible to test the validity of the model by comparison with empirical data

- it must be simple (if it is not simpler than the real situation, it is not a model)

- it must be manageable mathematically or on a computer

- it should ideally include all essential features of the modelled system and exclude non-essential ones.

1.2 Cont. VS discrete growth models (M1.1,2.1)

<u>Growth model</u>: mathematical model for the time evolution of a single species with $N(t)$ individuals at time t. Negative values of $N(t)$ are not allowed, the species become extinct if $N(t) < 1$ at any time. As a starting point, neglect spatial dependence, all interaction and competition for resources with other species, and external influences (most suitable for laboratory conditions, but can also give qualitative information about real-world species). To model how N changes with t we must consider whether generations overlap or not. Many species have little or no overlap between successive generations:

- A natural time for a new generation to appear is often one year.

- Some insects have new generations on the time scale of days.

- The time for cells can be hours.

- Bacteria and viruses can have shorter time between generations.

If there is a large overlap between generations or if the time between generations is small compared to all other time scales in the system, $N(t)$ changes smoothly in time and it is suitable to model its time evolution using a <u>continuous model</u>.

   If the overlap is not large, it may be more suitable to use a <u>discrete model</u>. Divide time t by the typical time between two generations to obtain discrete dimensionless time steps, $\tau = 0, 1, 2, \ldots$, such that N_τ denotes the population size at generation τ.

   Rate of change of population (conservation condition):

$$\Delta N = \text{births} - \text{deaths} + \text{migration} \, .$$

<u>The time evolution of N becomes:</u>

Continuous	**Discrete**
Instantaneous change	Change from one generation to next
ΔN is rate of change per time unit	ΔN is change per generation
$\dot{N} = \underbrace{\Delta N}_{\text{flow}}$	$N_{\tau+1} = \underbrace{N_\tau + \Delta N}_{map}$
$\Rightarrow$ Smooth <u>trajectory</u>:	$\Rightarrow$ Sequence of points (<u>orbit</u>):
$N(t)$ starting from $N(0) \equiv N_0$	$N_0, N_1, N_2, \ldots$

These are examples of continuous and discrete <u>dynamical systems</u> with the constraints $N(t) \geq 0$ and $N_\tau \geq 0$ (biologically relevant solutions). In one dimension the continuous system can be analytically solved in general. The discrete system has known solutions if ΔN is linear in N_τ, while non-linear ΔN rarely have analytical solutions.

1.2.1 Simplest example: Malthus model

<u>Linear model</u>: The change in population is proportional to size N with per capita birth rate b and death rate d (neglect migration)

$$\Delta N = bN - dN \,, \text{ with } b > 0 \text{ and } d > 0 \,.$$

Continuous	Discrete
$\dot{N} = bN - dN$	$N_{\tau+1} = N_\tau + bN_\tau - dN_\tau$
$\Rightarrow N(t) = N_0 e^{rt}$	$\Rightarrow N_\tau = N_0(1+r)^\tau = N_0 e^{\tau \ln(1+r)}$

With initial condition $N(0) = N_0$ and $r = b - d$ is the continuous per capita growth rate (often simply denoted 'growth rate' or 'birth rate'). Note that r is a rate (unit of inverse time) in continuous models, while it is dimensionless (per generation) in discrete models. Both cases show exponential growth ($r > 0$) or decay ($r < 0$), but with different rates: r (continuous) and $\ln(1+r)$ (discrete) [to compare the two we assume that r and t are measured in units of generations]. Which rate is correct depends on the system we want to model: if generations are overlapping the continuous result is more likely to apply, while if generations are non-overlapping then the discrete result is more likely to apply. If r is small then $\ln(1 + r) \sim r$ and the results of the continuous and discrete models coincide (in this limit the population does not change much per time unit and thus discretization of the continuous system is well approximated by the discrete system).

For the Malthus model the qualitative behaviour is the same (exponential growth) in the discrete and continuous system. As we shall see in Lecture 2, for non-linear systems the qualitative behaviour can be completely different in continuous and discrete systems.

1.3 Example: Growth model for harvesting (M1.6)

The following example illustrates how a simple model can be used to find interesting and important points about a complex biological system. Harvesting of renewable resources such as fish, plants, forest, or bacteria in a bio-reactor, requires a sustainable strategy. We typically want to ==find a strategy that gives a maximum long-term yield with a minimum effort and possibly with the constraint to affect the ecosystem as little as possible==. Consider the mathematical model

$$\dot{N} = \underbrace{rN\left(1 - \frac{N}{K}\right)}_{\text{Logistic growth}} - \underbrace{EN}_{\text{harvesting}} . \tag{1}$$

Assume $r, K > 0$ and $E \geq 0$. We remove population proportional to the number of individuals: the harvest yield is $Y \equiv EN$ with ==E being the harvest rate== ('effort'). In the absence of harvest ($E = 0$) the system (1) has a per capita growth rate $\dot{N}/N$ that depends on N:

$$\frac{\dot{N}}{N} = r\left(1 - \frac{N}{K}\right) .$$

For small values of N this rate equals r and we obtain exponential growth as in the Malthus model in Section 1.2.1. To take into account the effects of limited resources or overcrowding, the growth rate decreases linearly with growing N until it reaches zero at the ==carrying capacity== $N = K$. For values of $N > K$ the per capita growth rate is negative and the population decreases (births+deaths+ migration < 0). This choice of $\dot{N}/N$ decreasing linearly with N defines the logistic model (Verhulst model). Other growth models have other dependencies on N in their per capita growth rate.

Solving $\dot{N} = 0$ in Eq. (1) gives the steady states (fixed points):

$$N_1^* = 0 , \; N_2^* = K\left(1 - \frac{E}{r}\right) .$$

Visualization of the flow (1) determines the stability of the fixed points:
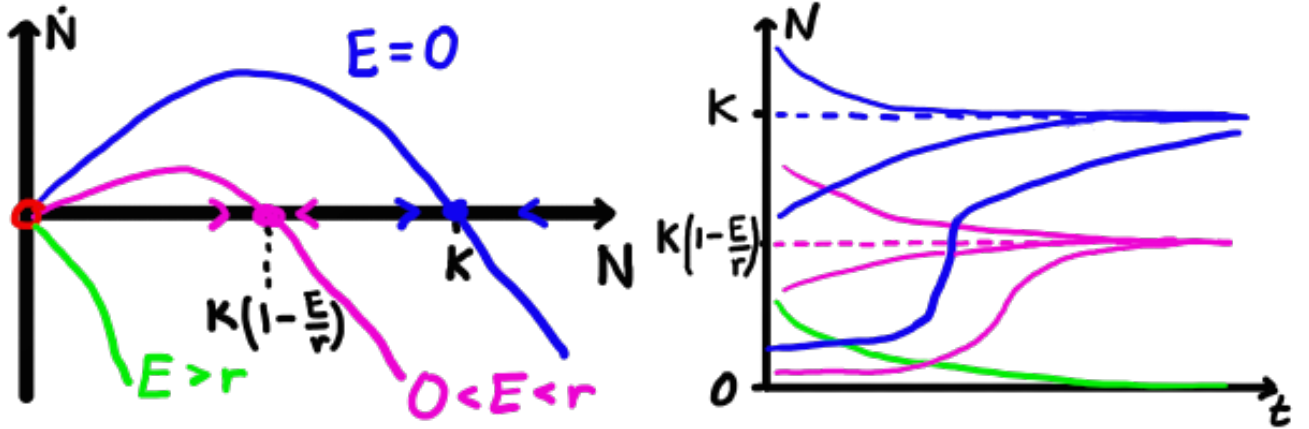
Example trajectories are shown in the right panel. If $0 \leq E < r$, the population eventually ends up at the positive, attracting fixed point N_2^*. For the case without harvest ($E = 0$), this steady state is equal to the carrying capacity, $N_2^* = K$. If the harvest rate is too high, $E \geq r$, the fixed point $N_1^* = 0$ is attracting and the population eventually dies out. Solutions starting close to a stable fixed point approach it exponentially fast, while, if N_1^* is unstable, solutions starting close to zero initially grow exponentially, giving a sigmoidal shape of the solution.

   The yield in the steady state is then

$$Y(E) = EN_2^* = EK\left(1 - \frac{E}{r}\right). \tag{2}$$

By solving $Y'(E) = 0 \Rightarrow E = r/2$ we obtain a maximal yield $Y_{\max} = Kr/4$ with population $N_2^* = K/2$.

   This analysis shows that to obtain a maximal yield we should harvest at a rate $E = r/2$, but it does not say anything about how sensitive this strategy is against perturbations coming from stochastic fluctuations of the population or from neglected factors in the model.

   One aspect not considered so far is the recovery time T_{R}, i.e. the time to move back to the steady state after a perturbation. We know that N_2^* is a stable steady state, the recovery time quantifies **how** stable it is. Write the system (1) as $\dot{N} = f(N)$ with $f(N) = rN\left(1 - \frac{N}{K}\right) - EN$. Taylor expansion around N_2^* gives

$$\dot{N} \approx \underbrace{f(N_2^*)}_{=0} + \underbrace{f'(N_2^*)}_{\lambda}(N - N_2^*) \;\Rightarrow\; N = N_2^* + \text{const.} \cdot e^{\lambda t}$$

close to N_2^*. Here $\lambda \equiv f'(N_2^*)$ is <u>stability exponent</u>, it evaluates to $\lambda = E - r < 0$ (fixed point is stable $\Leftrightarrow \lambda < 0$). Estimate the recovery time by the characteristic time scale close to the fixed point:

$$T_{\mathrm{R}}(E) \equiv \frac{1}{|\lambda|} = \frac{1}{r - E}.$$

To reduce the number of parameters, we choose to measure population in units of K, $N' = N/K$ and time in units of $1/r$, $t' = rt$:

Steady state population size $N_2'^* = \frac{N_2^*}{K} = 1 - \frac{E}{r}$

Steady state yield $Y' = \frac{Y}{rK} = \frac{E}{r}\left(1 - \frac{E}{r}\right)$

Recovery time $T_{\mathrm{R}}' = rT_{\mathrm{R}} = \frac{1}{1 - E/r}$

In these units everything is determined by the dimensionless parameter combination E/r:
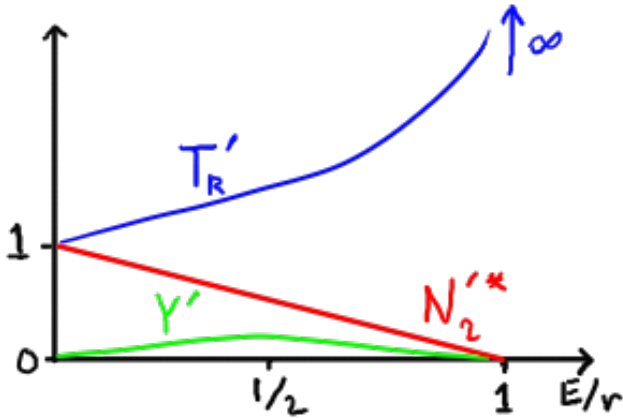
The steady state approaches zero as E/r approaches 1. At the same time the recovery time approaches infinity. The yield reaches its maximum at $E/r = 1/2$ and thereafter declines.

Within this model, the best harvest strategy is to have a harvest rate E=r/2, giving the maximal yield. Further increase of E is counterproductive: it decreases the yield and may eventually lead to catastrophic consequences: the population may become extinct or need a long time to recover even if we stop harvesting (c.f. the sigmoidal shape of N plotted against t with $E = 0$ and small N_0 in the right panel of the first figure in this section). In conclusion, the simple example model (1) shows that it is possible to approach catastrophic outcomes if we mindlessly increase the harvest rate beyond the maximal sustainable yield when harvesting renewable resources.

Another, even more catastrophic approach to harvesting, is to have a constant yield Y_0 (e.g. a constant fishing quota). The model becomes

$$\dot{N} = rN\left(1 - \frac{N}{K}\right) - Y_0\,.$$

Analysing this model shows that the time to recovery is

$$T_{\mathrm{R}}(Y_0) = \frac{1}{r\sqrt{1 - Y_0/Y_{\mathrm{max}}}}\,.$$

This model is very sensitive close to the maximal yield: as $Y_0 \to Y_{\mathrm{max}}$ the recovery time becomes infinite (the fixed point becomes marginal) $\Rightarrow$ this is not a good harvest strategy.

1.4 Example: Insect outbreak model (M1.2)

The following model for the population of spruce budworms subject to predation by birds illustrates hysteresis shown in many biological populations or ecosystems

$$\dot{N} = rN\left(1 - \frac{N}{K}\right) - \underbrace{\frac{BN^2}{A^2 + N^2}}_{\equiv p(N)\,,\ \text{predation}}\,.$$
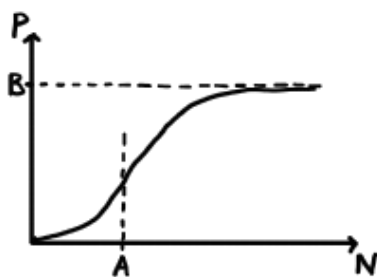
When N is small, there is no predation (birds seek food elsewhere). When the number of budworms is large, the effects of predation saturates (birds eat as fast as they can). This system shows outbreaks and hysteresis as the carrying capacity is slowly changed (details in Murray 1.2).

If you need more training, you can have a look at the training question on an insect outbreak model. The solution will be presented as part of the first problem session.