Computational Biology Course Exercises

2024

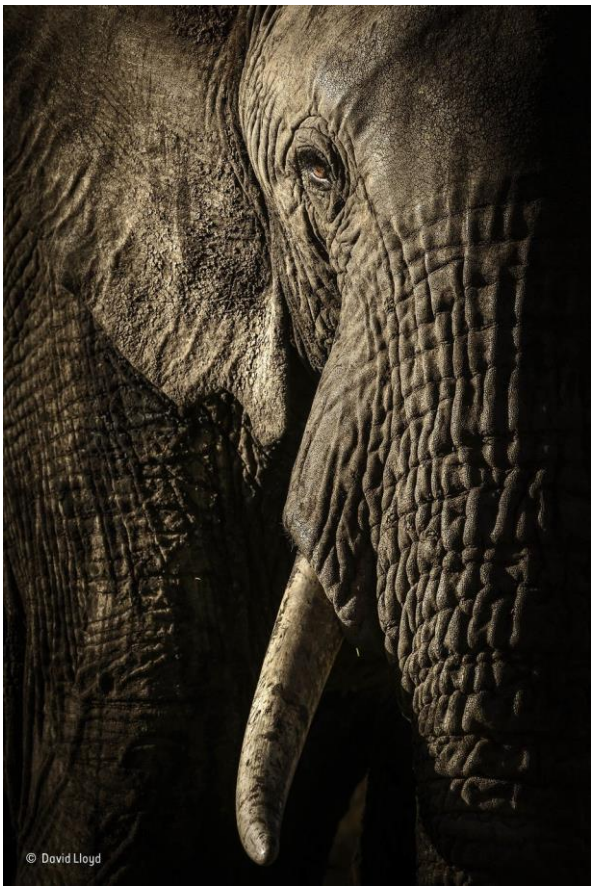
READ THIS FIRST

This document contains exercises to study simple biotic systems using different model formalisms (mainly differential equations and cellular automata).

NOTE! While doing these exercises, keep in mind that...:

- ... the questions are embedded in the broader context of the course / lectures
- ... formulating a clear bottom line at the end of each exercise is crucial
- ... knowing the answer to a question does not necessarily mean you fully understood it

The questions will be discussed during the Monday afternoon review sessions



"Elephant, I ..."

Course introduction

In the course Computational Biology, you will learn:

- How (not) to use modeling to gain insights into biological systems
- Biological insights obtained through modeling

Where to find course information:

Most communication will happen via the course website http://theory.bio.uu.nl/BINF. Here you will find the time table, sheets, software, references, and the course reader. Note that the course reader found there should **not** be seen as a good alternative for attending the lectures, but should instead be used as a self-study tool to complement your understanding. The website also contains hyperlinks to recorded lectures from 2014-2015 and last year. Because of COVID-19, it is likely that there will be a hybrid format and lectures can be recorded. However, you should only follow the lectures from home if you need to quarantine.

Rough time schedule:

The course is roughly split into two parts. The first part, which is covered by the first one or two weeks of the course (both lectures and exercises), will mostly focus on modeling formalisms and how to use them. In the weeks after that we will discuss what theory has been obtained through modeling with a great variety of topics, ranging from the origin of life, evolution of robustness and evolvability to group behaviour, development, metabolism and much more. This part serves to show how the modeling formalisms from the first part can be applied in an interdisciplinary way, and promotes a critical view of how to use models (and how NOT to!) to gain insights into biological systems.

You have six weeks for (at least the first 7 (0-7)) exercises contained in this document. After that, we switch to mini-projects, which you will work on for two weeks, culminating in a report and short presentation. During this mini-project you will make simple models yourself and analyse them. After this, there is one week in which you work on and hold a literature seminar, where we critically discuss and integrate recent literature in the context of what you've learned in the course. The final week is for studying for the exam.

Examination:

On Thursday of the last week, you will do a written exam (on the computer, counts for 85% of your final grade). In order to get a final evaluation your mini-projects and presentation at the literature seminar need to be satisfactory (>= 5.5), together these count for 15% of your final grade, 7.5% each.

Working on your Laptop or computers in CLZ:

During the practicals you can either work on your own laptop, or on computers in the CLZ. Your laptop needs to be running Ubuntu (or other forms of Linux) or MacOS for this course. Windows users can try using 'Ubuntu for Windows'. Alternatively, you can work **in Ubuntu** on a PC in the CLZ (practica rooms with computers at the UU) by logging in there and running two commands. To install the required dependencies and set all this up on each system (Linux, Windows, MacOS or Linux on a UU computer) check the manual (https://tbb.bio.uu.nl/BINF/software/README_FIRST_InstallingDependenciesCourseSoftware.pdf). You will only need Ubuntu for exercise 7.

Cellular automata modelling environment:

We use the online framework CACATOO (https://github.com/bramvandijk88/cacatoo), which is based on javascript and allows you to make and edit interactive (spatial) models easily and with high speed in the browser. It is advisable to **make an account on JSFiddle (https://jsfiddle.net/)** so you can easily clone example code and save it to your account. Be sure to refer to the CACATOO documentation (https://bramvandijk88.github.io/cacatoo/overview.html) when you want to do something specific in the code but don't know how (such as getting a random Neighbour, assigning colours, plotting a graph, etc.).

0 Introduction to Cellular Automata

0.1 Game of Life & Modulo Prime

Examine the behaviour of two example cellular automata: Game of Life (https://jsfiddle.net/DieStok/sdvw10j9/) and Modulo Prime (https://jsfiddle.net/DieStok/ez32mnau/). Because these CAs update extremely fast, they are paused by default and you need to unpause them. Buttons are provided to control the speed of updates, as well as initial conditions and to draw your own patterns if you want. **Before you start, make sure to clone the JSFiddle, so what you are changing is saved!** Continue the runs for a while and carefully observe what happens. Study the behaviour of both systems on different space and time scales and answer the following questions for each model in sequence.

a Verbalize what you see.

b "Mesoscale pattern formation" occurs.

- Explain the term "mesoscale".
- Distinguish different types of mesoscale patterns in terms of their spatiotemporal dynamics.

c Characterize of the observed dynamics in terms of attractors in the full state-space, or in subspaces thereof.

d How do the dynamics depend on the boundary conditions and initial conditions?

e Compare your observations with the properties of the CA rules as discussed in the lecture. Focus on "predictability".

f Examine the effect of asynchronicity. For this, you will need to change the update rule in the code.

0.2 Voting rule

A voting rule is a next state function for which the state at time $t + 1$ depends on the number of neighbouring states at time t.

a Consider a cellular automaton with two states: 0 and 1. Investigate voting rules in which the next state is 1 if the sum of the neighbouring states > n. An example is Vote (https://jsfiddle.net/DieStok/yz5huc49/), which is a CA where 9 neighbors (self-inclusive) are taken into account. Investigate the CA when n is set to 4 (meaning that the next state is simply the majority of neighbouring states). What happens for different values of n?

b A stochastic automaton is an automaton in which errors are introduced according to a certain probability distribution. To program such an automaton the function genrand real1() can be used. Choose $n = 4$. Compare the behaviour of a deterministic

and a stochastic automaton (i.e. one in which a grid point has a chance to not follow the majority).

c Slightly adjust the voting rules by assuming that still five different inputs result in next state 1, and the five other inputs yield next state 0, but that the 10 different input-output pairs are not strictly ascending in terms of input values (not all the inputs resulting in next state 0 have smaller values than the inputs resulting in next state 1, e.g. swap the outcomes for 4 and 5 neighbours). Study the behaviour of these adjusted voting rules, and compare this with your previous observations.

d Compare the behaviour of a deterministic synchronous and an asynchronous automaton.

e Investigate the behaviour of the voting rule in a well-mixed system.

f Give examples to which all these cellular automata (Game of Life, Modulo Prime, Vote) could apply (processes they could be caricatures of). If you can't or think there are none, explain why.

0.3 Structural stability

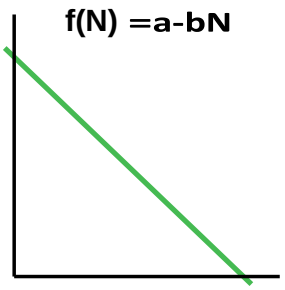
A model is considered to be **structurally stable** if a small change in the model formulation does not cause major changes in the outcome of the model, like what you have seen above with stochasticity and the timing regime of the models.

a To what extent do the models studied in Exercises 0.1 and 0.2 exhibit "structural stability"?

b Is structural stability a desirable property of a model? Discuss the implications of structural stability for modeling.

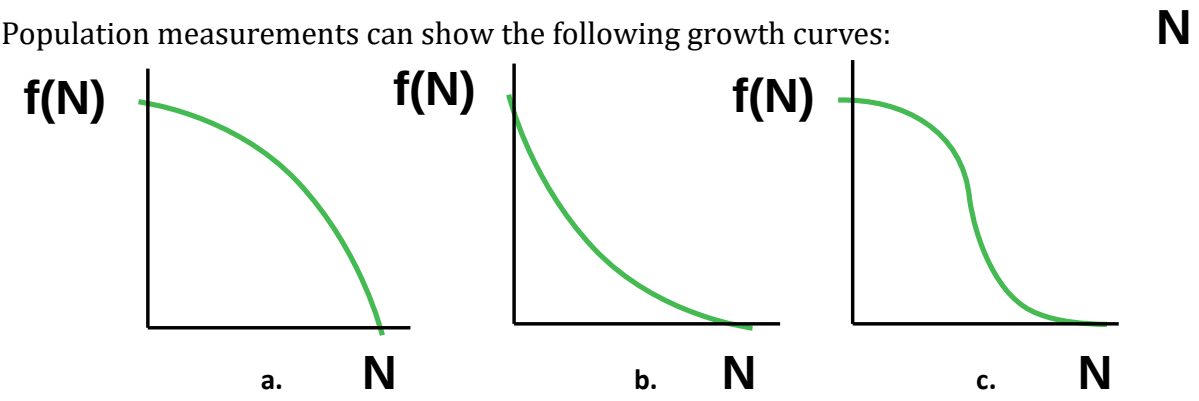
! NOTE: Answering this question also in later exercises may be useful !

1 Classical theory of growth: "Unique next state function"



Population growth is often described by the logistic equation: $\frac{dN}{dt} = f(N) \cdot rN$, where $f(N)$ is a function that describes the growth rate per individual, and r is the intrinsic growth rate. In the case of the logistic equation, $f(N)$ is defined by $f(N) = a - bN$, where a and b are positive constants.

Population measurements can show the following growth curves:



a Compare these measurements with the often used $f(N) = a - bN$, and identify the different biological processes and assumptions captured by all 4 cases. Which (implicit) assumptions do they have in common?

b Compose two approximate, simple ODE models for curves a and b. Study their behaviour in grindR, an R package available here: http://theory.bio.uu.nl/rdb/grind.html. (e.g. study how an initially small population grows to carrying capacity for both functions)

c Study the occurrence of chaotic behavior in both models (a and b) and the original logistic equation in a difference equation (MAP). You can use the *logisticMap.R* file available on the website as a starting point: https://tbb.bio.uu.nl/rdb/grindR/logisticMap.R.

- Is the intrinsic growth rate r a sufficient predictor for chaotic behavior?

d Compose a very simple model of a growing (and dying) population using a stochastic CA. Instead of sim.initialGrid(...) you can use the function sim.initialSpot(sim.YourGridNameHere, 'YourStateNameHere', 1, 2, sim.YourGridNameHere.nr / 2, sim.YourGridNameHere.nc / 2) to initialise a circle of 2 individuals in state 1 in plane YourStateNameHere, in the middle of the grid YourGridNameHere. Run your CA and observe.

- Is there density dependence?
- Did you define density dependence explicitly?

e Write an ODE that corresponds to your CA as a mean field approximation.

- Is there a term that expresses density dependence?
- Does your mean field approximation correspond to any of the equations above?

f In the CA from before, draw $f(N)$ against N for this growth model (check *Counting total populations* and the function sim.YourGridNameHere.PlotXY() in the

CACATOO manual: https://tbb.bio.uu.nl/BINF/software/FunctionReferenceCACATOO.pdf . Which of the curves above would describe the dynamics? Reconsider question **a**.

g Let us suppose that we are to measure $f(N)$ of real organisms.

I How would you measure $f(N)$ if you had relatively little time to perform your experiment (i.e. you cannot wait for more than a few generations).

II Do you expect to obtain the same graph as that obtained in **f**?

III What kind of (computer) experiments can you imagine that could obtain the same result as in **I** with one simulation run of the CA model? Try it.

h Discuss whether or not the previous findings (regarding N and $f(N)$) contradict the concept of a "unique next state function", i.e. given the current population size (N) of the system, the next state is uniquely determined.

i Study the growth of an initially small, local population with 8 evolutionary neutral strains. Go here (https://jsfiddle.net/DieStok/271h954u/). Formulate a biological conclusion from the growth pattern that you see.

j This program (https://jsfiddle.net/bramvandijk88/0mu9ah5p/) is a CA model for diffusion limited growth: the species in this model can grow when encountering diffusing food particles. Study how this mode of growing affects the **population growth** function and compare it to those you found in **f**.

k Compare the dynamical behaviour of the CA from question **f** with that found in **c** (consider both micro- and macro-scale chaos).

2 Modeling formalisms, state space and the cell cycle

First read Li et al. (2004) [1]

The authors comment on the "surprising robustness" of the cell cycle regulatory system.

a Distinguish at least three notions of 'robustness' used in the model.

b How does the cell cycle network differ from similar random networks for each of these types of robustness?

c Compare their results qualitatively with what is known from random boolean networks (Kauffman networks).

d From what you know about biological transcription regulation networks, do you see a limitation of the threshold function used that Kaufmann boolean networks do not have? What biological regulating mechanism do you have in mind?

e Can you determine from the figure below (Figure 1) whether they use synchronous or asynchronous updating of the network ? How?

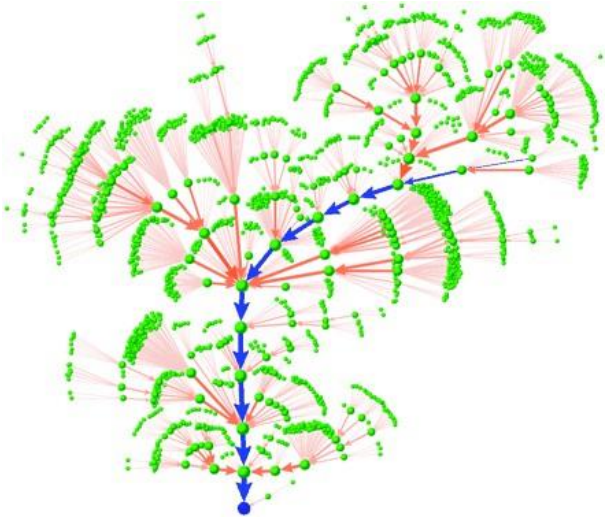


Figure 1: State space of yeast cell cycle network

The following has been proposed as a VERY SIMPLE phenomenological model for the cytokine regulation of the cell cycle:

$$\frac{dX}{dt} = k_3(Y - X)(a + X^2) - k_2X \qquad \frac{dY}{dt} = \varepsilon(k_1 - k_2X), \quad \text{where } a, k_1 \ll k_2, k_3 .$$

f Study the model (make the dynamics of *Y* much slower than those of *X*) and analyse it in terms of nullclines and vector field (feel free to use grind.R). Tune the parameters such that the model qualitatively behaves like the cell cycle network

[1] All the references mentioned in the excercises are available from http://bioinformatics.bio.uu.nl/BINF/References for_exercises/

above. Compare Figure 1 with the state space figure you have just come up with. Identify similarities and differences between the two pictures.

g Discuss the robustness of the 2 variable ODE model.

h The cell cycle involves many processes such as DNA replication, spindle formation, cell growth, etc. Furthermore, the expression levels of a large proportion of the genes is changed during the cell cycle. Explain how these downstream effects could be regulated in the two models and discuss the robustness of this regulation in either case.

i Embryonic cells apparently do not have a volume-induced 'start' mechanism, but continuously divide anyway. Think how both models could be modified to incorporate this observation.

3 Classical Ecological Models and Quasispecies Theory

3.1 Competition and the voting rule

a Compose the simplest possible CA of 2 competing species (for example, 1 and 2 are species and 0 is empty space in which they reproduce).

b The following is a system of ODEs that describe two competing species:

$$\begin{aligned} \frac{dX}{dt} &= a_1 X - b_1 X^2 - c_1 XY \\ \frac{dY}{dt} &= a_2 Y - b_2 Y^2 - c_2 XY. \end{aligned}$$

Investigate the five different possible outcomes of this system (use nullcline-analysis with pencil & paper only) and give a biological interpretation for each of the configurations.

c Which of these five possibilities have you implemented in **a** above? Can you adapt that CA to make a correspondence with the cases found in **b**? Compare the behaviour of the CA models with that of the ODE system. To what extent are the dynamics of the CA models the same as those of the ODE model? Why/why not?

d Compare these CAs with the voting rule system. Which of the five models would agree best with the voting rule CA? What do you expect to be the influence of regular disruptions (e.g. hurricanes, raging fires, etc.) on the system?

E. coli can harbor plasmids which can produce many different types of colicins. Colicins are proteins that kill *E. coli*, unless the bacterium possesses a specific resistance factor against it. Colicins and their corresponding resistance factors are genetically coupled, both coded on one plasmid. The bacterium is resistant against colicins because it does not express certain membrane proteins, which normally play a role in nutrient uptake. Therefore, production of colicins and resistance against them have a negative effect on the growth of the bacteria.

e Parameterize the ODE model above that you studied in **b** based on the above description, and study it.

f Is the colicin-producing strain able to invade?

g Experiments show that a colicin-producing strain inoculated onto an agar plate can quickly overgrow a present sensitive culture. Explain this. Verify your answer by testing this in your CA from question **a**.

3.2 Predator-prey dynamics and structural stability

"... Sebastiaan is opgeveegd.", Annie M.G. Schmidt.

Humans sweep spiders. Spiders eat mosquitoes. Such predator-prey interactions have been studied using the classical Lotka-Volterra predation model:

$$\frac{dM}{dt} = aM - bSM$$
$$\frac{dS}{dt} = bSM - dS$$

a Use the classical Lotka-Volterra predation model to show that sweeping spiders will not reduce the number of spiders in one's house (use pencil & paper; if needed, check with grindR). Consider two different ways of sweeping spiders.

b Is the conclusion about the spiders an artifact of the model? For example, is the result sensitive to small changes in the model structure?

c Is it possible to compose a simple CA model with identical (explicit/implicit) assumptions as the classical Lotka-Volterra model? Which implicit assumption is made in the CA? Check by writing a mean field approximation of your CA.

d Study the effect of killing spiders and/or mosquitoes in a CA and discuss what is happening on multiple scales. *(tip: also look at space-time plots, which were used earlier in Vote; It might also be nice to make a zoom-in plot to really see what is happening, for this, add in sim.createDisplay("YourGridName", "YourStateName", "(zoom in on top-left)", 30, 30, 10)*).

e Comment on the concept "structural stability" of a model.

3.3 Classical theory of prebiotic evolution

a Write down a replicator equation, replacing the chemo-stat assumption by competition for a resource (e.g. space). Use the simplification into two variables; write one equation for the master sequence and one for all the other sequences.

b Study the model in terms of nullclines for different mutation rates and selection coefficients. Sketch the nullclines of the model and use grind.R to check your sketch. Calculate the conditions for co-occurrence of both the master and the mutants (use pencil and paper). Connect this with the error threshold.

c Study the error threshold in a CA model. Compare the error threshold you find here quantitatively with what you find for the mean-field approximations. Give an explanation for your observation. Is the difference in the error threshold due to the stochastic nature of the CA or due to something else? Check it!

4 Competition and speciation

4.1 Sympatric speciation

Read Dieckmann and Doebeli (1999). For this question, focus on the *clonal cases* rather than the cases with *sexual reproduction*.

a Figure out the correspondence between the model of the paper and the ODE model of Exercise 3.1. Draw a cartoon strip in nullcline pictures of the sympatric speciation as described in Fig. 1a of Dieckmann and Doebeli (1999) starting with low K. (Hint: start with the resident and an identical mutant, and slowly change the mutant.)

b Rephrase the speciation condition $\sigma_c < \sigma_K$ in terms of (the change of) competition parameters. How can you see this in the cartoon?

c Translate the same condition into a verbal characterisation of the feeding habits of the species involved.

The paper is on *sympatric* speciation, i.e. on speciation in well-mixed systems. Consider the equivalent model in a CA (It is not necessary to make the model, but answer the following questions based on the insights from earlier exercises).

d Will the 'parapatric' situation (not well-mixed situation) help or hinder the speciation? In what sense would you still call the resulting speciation, if it happens, 'sympatric'?

4.2 Ecosystem diversity

a Consider the Exercises 3.3 and 4.1. Is the process studied in each exercise speciation? Why/Why not?

b Compare the model in Exercise 3.3 with that in Exercise 4.1 in terms of the relationship between evolutionary dynamics and population dynamics.

c Comment on "competitive exclusion" and "survival of the fittest".

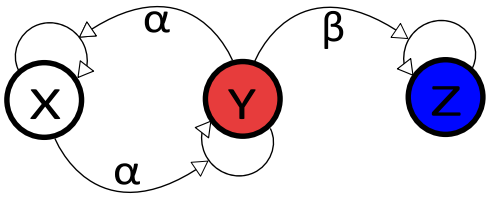
d How does the consideration of space influence our understanding of ecosystem diversity? Consider the results of Exercise **3** and **4.1**. Make a bottom line.

5 Short term versus long term advantage

"Honesty pays ... ?"

5.1 Mutualists and a cheater

Consider the following interaction structure in which the interaction strengths between both obligate mutualists are equal. Note: **there is no independent self-reproduction, we assume that self-replication is only catalysed with the rates α and β.** We compare an ODE model with CA models with similar assumptions.



a Construct an ODE model, and study it. Is it possible that all three species coexist if α 6= β?

b Consider the following model expressed in CACATOO code. Make a correspondence between this model and the above system. What are α and β in this model? To what extent are the processes considered the same as in your ODE?

BEGIN

```
let sim;
var A2B = 1.0        // Mutualist species A giving help to reproduce B var B2A = 1.0       // Mutualist species B giving help to
reproduce A var B2C = 0.9      // Mutualist species B giving help to reproduce C ("cheater") var stay_empty = 1.0           //
Constant which scales the probability that nothing happens when
//competing for empty grid point, iow "stay empty"
var death = 0.1                                    // Death rate of individuals
let mdif_interval = 0

function cacatoo() { let config = { title: "Mutualists and cheaters", description:
    "", maxtime: 1000000, ncol: 100, nrow: 100, // dimensions of the grid to
    build wrap: [true, true],    // Wrap boundary [COLS, ROWS] seed: 56,
        fps: 60,   // Note: FPS can only be set in fastmode fastmode: false, scale:3, // scale of the
        grid (nxn pixels per grid point) graph_interval: 10, graph_update: 50, statecolours: {
            'species': {
                1: "#FFFFFF",
                2: "red",
                3: "#3030ff"
            }
        } } sim = new Simulation(config) sim.makeGridmodel("cheater")
    sim.initialGrid(sim.cheater, 'species', 1, 0.33, 2, 0.33, 3, 0.33)
// Place the three 'species' in grid points (33\% A, 33\% B, 33\% C) sim.createDisplay("cheater", "species", "Mutualists and
    cheater")

    sim.cheater.nextState = function (i, j) { let state =
        this.grid[i][j].species;
          if (state == 0)                                                    // If there is no species here
        { sumA = this.countMoore8(this, i, j, 'species',1); // Count the number of species 1 (mutualist A) sumB =
            this.countMoore8(this, i, j, 'species',2);          // Count the number of species 2 (mutualist B) sumC =
            this.countMoore8(this, i, j, 'species',3);          // Count the number of species 3 (mutualist C)

            pA = (B2A * sumB) * sumA;          // Relative chance that A wins pB = (A2B * sumA) *
            sumB;          // Relative chance that B wins pC = (B2C * sumB) * sumC;          // Relative
            chance that C wins psum = pA + pB + pC + stay_empty;
// Total = pA+pB+pC+stay_empty (scales the chance that nothing happens during competition) ran = this.rng.random();
          // Draw a single random number which decides 1 winner from "roulette wheel" (see below)

            if (ran < pA / psum) this.grid[i][j].species = 1 else if (ran < (pA + pB) / psum)
            this.grid[i][j].species = 2 else if (ran < (pA + pB + pC) / psum) this.grid[i][j].species = 3
        //else: (no winner, spot stays empty for now)
        }

        if (this.rng.random() < death)
// Stochastic death (species become 0, which is an empty space for the next step to compete over) this.grid[i][j].species = 0
    }

    sim.cheater.update = function () { this.synchronous()        // Update all grid points based on the next-state function (defined above) if
        (this.time % mdif_interval == 0) this.MargolusDiffusion() // Every so often mix individuals a bit this.updateGraphs()        // OPTIONAL:
        add some graphs
    }
```

BEGIN

c Is it possible for three species to coexist in the model of question **b**? If coexistence is possible, what are the conditions (i.e. possible values of α and β?)? Explain the results. Use this implementation of the model (https://jsfiddle.net/DieStok/dz86jLcu/).

d Can the ODE model of question **a** be a good mean-field approximation of the model of question **b** in certain conditions? How can you examine this? Try it.

e To examine the results of question **d**, compose another CA model, which behaves in the same way as the ODE model if the system is well-mixed, but also behaves in a similar manner to the model of **b** if the system is not well-mixed. [Hint: Compare the competition between the two mutualists and a mutualist and the cheater in both models. Consider indirect vs. direct local interactions.]

f Comment on the time scale of local interactions in the two CA models.

g Without changing the 'NextState()' of question e, can you adapt the code in such a way that it behaves the same on in the long run as the well-mixed version of the model of question **b**?

h Distinguish the effects of local interactions from the effects of pattern formation on the behaviour of a biological system.

i Go here to see a model where the above species can evolve how much help they **get** (their respective alpha and beta, with no upper limit). Who "out-evolves" whom? **optional extra:** What do you expect will happen if X has a cheater of its own? (try it if you're interested!)

5.2 Ants and ant plant

Quickly scan Yu and Pierce (1997). How is it possible that Azteca and Allomerus coexist? Comment on the requirements for coexistence suggested by the authors.

6 Multilevel selection

6.1 Ageing bacteria

In 2011, the article "Bacteri¨en worden ouder en juist dat houdt de populatie fit" (*Bacteria age and that is what keeps the population fit*) was a news item on the website nu.nl. • Read the article (you can find it under 'references' on the course website). For those who need it, a translation is provided. Can you interpret the article in terms of a mechanism discussed in the lectures?

6.2 CPM

Multilevel selection arises automatically in spatially embedded replicator systems, where emerging self-organised patterns become a new level of selection.

Here we examine systems where two levels are prior defined: a cell level and a within cell level. The cell level is defined through the CPM modeling framework, whereas the within cell dynamics is defined as a map (difference equation). Two cell types can be used to design competition experiments. You can find an online implementation of StochasticCorrector here: https://jsfiddle.net/alkminion/dufsy5a2/1/ .

a Examine the model. Run the program with default parameters. Click on "Show Y molecules". Verbalise what you see. Why do the cells eventually die?

b How can cells survive by multilevel selection, when they CANNOT without multilevel selection? (cf lectures and also take the article from exercise 6.1 into consideration). Change the parameters in such a way to check if (and when) this works.

c How does the mechanism in **b** depend on the timescales of the different levels? Which parameter of the cell-level can you change to tune the timescales without changing the speed of the internal dynamics? Try it.

d In the model, you can also add a second cell-type by setting NRCELLS[1] to start with a few cells. How does the mechanism described in **b** influence the outcome between two cells? If the internal dynamics were not important for cell growth, would the outcome of competition be the same?

e Can two cells with the same (instable) internal dynamics coexist? Why / why not? (cf think also of the coexistence as we studied in earlier exercises)

f Revisit the error threshold in this model.

6.3 Ageing bacteria, revisited

a After answering the previous questions, scan Rang et al. (2011) on which the news item on bacterial aging is based. To what extent does your previous interpretation

make sense? Is there a conflict between the mechanism discussed by Rang et al. (2011) and the mechanism you studied in 6.2b?

7 RNA evolution: topology vs. energy optimization

In the lectures we have seen a number of properties of the evolution of RNA secondary structure. The redundancy of the genotype-to-phenotype mapping has significant consequences for the information threshold. To gain a better understanding of the dynamics of RNA evolution, you are going to use the RNAevol and RNAlocal tools that are available on the course website, together with their manual[2]. Note that here, the Linux environment or Linux subsystem for Windows (e.g. Ubuntu) is absolutely necessary, as you need to work in a terminal to start these programs using ./RNAevol and ./RNAlocal (after setting permissions as described in the Readme file).

a Evolve a homogeneous population to the given target structure.

- Explain what you see in the graphs.
- Explain how the evolution of Hamming distance to initial sequence relates to that of the Structure distance to the target structure. • When do peaks in the mean square displacement per timestep occur? Why? [Hint: look also at how the population changes.]
- Explain the behaviour of the sub-populations (of most abundant genotype, fittest and most abundant genotype and of fittest replicators) from the moment a fitter individual appears.

b What happens if you change mutation rate and/or population size? Look in particular at:

- timescales of the dynamics
- relative sizes of the sub-populations
- the relation between the fittest replicator and the average distance to target structure
- mean square displacement before and after a transition to a fitter individual

c Explain the changes in the graphs when you evolve toward a target sequence instead of a structure. (To do so, use the option "-Selection 3")

d Go back to evolving structures. Find the error threshold (think about initial conditions!).

e What features that you saw in the various graphs can you relate to the difference between the genotypic and the phenotypic error threshold? Consider both the local and the global fitness landscape. In other words: verbalise the conclusions from this question so far.

f Study how neutral evolution changes the *local* structure landscape (use one of the random sequences you can find on the course website). Use the script `locallandscape.bfile` to visualize the local landscape in xmgrace:

[2] Available from https://tbb.bio.uu.nl/BINF/software/RNA/

xmgrace -block filename outputRNAlocal -batch locallandscape.bfile. If you don't find what you might expect from the lectures, explain why that is and test your hypothesis.

Read Kun et al. (2005). The authors make an estimate of the fitness landscape of real ribozymes by combining mutagenesis data with structure predictions of mutants. Their estimated error threshold permits rather long genomes. We will examine the neutrality of the ribozyme sequence ourselves with our familiar tools.
The sequence and folded structure of the Neurospora VS ribozyme can be found on the course website.

g What λ do you find for the ribozyme, when considering only secondary structure as a fitness measure? Does the claim of the authors, that their fitness measure explains the relaxed Error Threshold, hold?

h Kun et al. found high values for the effective selection coefficient. Determine the selection coefficient of the sequences yourself in a way analogous to Kun. How does your answer relate to Kun's finding and to the predefined selection coefficient in RNAevol ($\sigma = 1.5^p$, where p is phenotypic distance)? What does this teach you about effective selection dependent on how far you are from the error threshold? [Hint: Take phenotypic distances of all sequences in the population (near the error threshold) into account.]

8 Diverse / miscellaneous modeling questions

8.1 Local interactions

Consider the following next state function of a cellular automaton, which is a simple model for a biological process:

```
var b = 0.2;
var q = 0.5;
var z = 0.01;
var r = 0.01;
var d = 0.05;

sim.model.nextState = function(i, j)
 {
  let rn = this.randomMoore8(this, i, j).val

  if (this.grid[i][j].val == 0) {
   if (rn == 1 | rn == 3) {
    if (this.rng.random() < b) {
     this.grid[i][j].val = 1
    }
   }
  } else {
   if (this.grid[i][j].val == 1) {
    if (rn == 2) {
     if (this.rng.random() < q) {
      this.grid[i][j].val = 2
     }
    }
   } else {
    if (this.grid[i][j].val == 2) {
     if (this.rng.random() < z) {
      this.grid[i][j].val = 3
     }
    } else {
     if (this.rng.random() < r) {
      this.grid[i][j].val = 1
     }
    }
   }
   if (this.rng.random() < d) {
    this.grid[i][j].val = 0
   }
  }
  }
```

a Give an interpretation of the biological process modelled. Describe what the parameters *z*, *r* and *d* stand for.

b Write down an ODE mean field approximation for the modeled interaction.

8.2 Modeling chemotaxis

When modeling diffusion, entities (e.g. cells) can be represented in CA-like models in three different ways:

1. One entity in each automaton (gridpoint)
2. Many entities in each automaton (gridpoint)
3. One entity in many automata (gridpoints)

a Formulate a diffusion model in each of the three representations, i.e. a model in which the entities move according to a diffusion process. You can use pseudocode.

b Explain for each case in what way your model deviates from CAs in the strict sense.

c Discuss particle conservation in each of the representations.

d How would you model two entities that diffuse with different speeds in each of the representations?

e Assume that there is an external gradient in the system (e.g. from "North" to "South"). Extend your models to include chemotaxis, i.e. a model in which the entities move dependent on this gradient.

There are different mechanisms of sensing a gradient:

1. Entities can directly observe the local value of the concentration and the direction of the gradient.
2. Entities can observe the local value of the concentration and have a short-term memory (i.e. remember the value sensed at time $(t - 1)$).
3. Entities can observe the local value of the concentration only.

f Discuss if and if so, how these sensing mechanisms could be formulated in the three representations mentioned above.

Required reading for the exercises

The following list of references is used in the exercises. It is recommended to print and read the articles cited in upcoming exercises, before the start of the practical.

References

U. Dieckmann and M. Doebeli. On the origin of species by sympatric speciation. Nature, 400:354–357, 1999. (Cited on page 11.)

A. Kun, M. Santos, and E. Szathmary. Real ribozymes suggest a relaxed error threshold. Nat. Genet., 37(9):1008–1011, Sep 2005. (Cited on page 18.)

F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang. The yeast cell-cycle network is robustly designed. Proc Natl Acad Sci, 101(14):4781–6, 2004. (Cited on page 7.)

nu.nl. Bacteriën verouderen toch wel. http://bioinformatics.bio.uu.nl/BINF/References_for_exercises/nu-nl-bacterien-verouderen-toch-wel-.html, 2011. [Online; accessed 28-October-2011]. (Cited on page 15.)

C. U. Rang, A. Y. Peng, and L. Chao. Temporal dynamics of bacterial aging and rejuvenation. Curr. Biol., 21(21):1813–1816, Nov 2011. (Cited on page 15.)

C. Yu and N. Pierce. A castration parasite of an ant-plant mutualism. Proc R Soc Lond B, 265:375–382, 1997. (Cited on page 14.)