Spreadsheets in Education (eJSiE)

Volume 10 | Issue 3 Article 3

2-27-2018

Mathematical Modelling and Computer Simulations in Undergraduate Biology Education

Gavin A. Buxton
Robert Morris University, buxton@rmu.edu

Mathematical Modelling and Computer Simulations in Undergraduate Biology Education

Abstract
A course in computational biology that introduces undergraduate biology students to mathematical modelling and computer simulations is described. Spreadsheets offer the perfect environment to introduce our biology students to computational thinking and the increasing role that computer simulations are playing in biology research. Here, we detail the spreadsheet modelling of some of the simulations covered in the course; the Lotka-Volterra predator-prey model, a cellular automaton model of tumor growth, and a model of an infectious disease outbreak. The experience of implementing computational biology simulations in a spreadsheet environment encourages and enables our biology students to use computer simulations and spreadsheets more in their future research, and makes our students more comfortable when interpreting scientific literature that pertains to computational biology research. These are important skills that our biology students will need in their future careers as researchers and scientists.

Keywords
Biology education; computer simulations; spreadsheets; finite-difference; cellular automata.

1 Introduction

Undergraduate biology education can be enhanced through the introduction of mathematical models and computer simulations by improving general quantitative skills, ensuring students are more comfortable interpreting the results from computational studies and encouraging students to implement computer models as part of their future research. There have been many studies that have linked the use of computational tools in STEM (science, technology, engineering and mathematics) classrooms to the enhanced learning of the students [1]. This is potentially more beneficial in biology classrooms as, in contrast to other STEM disciplines, biology majors are disproportionately composed of more female students [2]. Gender identity within our society often reinforces a perception that boys prefer computational, engineering or mathematical based careers while girls show a preference for the life sciences [3]. A potential gender gap in the use of mathematical models and computer simulations in future biomedical research could be alleviated through the use of a computerized learning environment, and by increasing the familiarity and proficiency of biology undergraduates with computer simulations in female-dominated classrooms [4]. In addition, there is strongly believed to be a technological gap between working and educated classes, the former including disproportionately many African Americans and other people of color, and bridging this gap at the university level is vitally important for promoting the diversity of future biology researchers [5, 6]. Given the important role that computer simulations will play in biological research (research in areas such as sustainable food growth, sustaining ecosystems and biodiversity, and understanding human health [7]) it is important that all of our students (regardless of race, gender or socioeconomics) are proficient in the use of mathematical models and their exploration with computational simulations, and better prepared to meet these future challenges [8]. This is in contrast to how mathematics, computer science and biology courses are currently taught. Mathematics is taught

almost entirely independent of undergraduate biology courses [9, 10], with students unable to appreciate the influences and overlap between these disciplines. Even more alarming is that biology majors often take few or no computer sciences classes. This has a direct and measurable impact on future biologists and their research, with papers containing a greater concentration of mathematical equations often receiving less citations by their peers [11, 10]. This lack of comfort that biologists have towards mathematically-dense papers limits their ability to appreciate the increasingly larger role that computer simulations are playing in biological research [10].

The majority of undergraduate biology textbooks are densely filled with facts and images, and lack insights into the nature of scientific inquiry [12] and the computational skills required for future biologists [10, 13, 14]. This is changing, however, with computer simulations increasingly being used in undergraduate biology courses [15]. It has been argued that computer simulations can allow students to perform experiments that might be too impractical, dangerous or expensive in reality [16, 17]. Furthermore, at the high school level there are concerns for animal welfare, and arguments have been made for replacing traditional dissections with computer simulations [17]. This often results in computer simulations being used as a black box tool for students to interact with on the short-term [17]. Some of the computational tools that are used for enhancing instruction, however, are both complicated and enlightening; for example, the Quantitative Circulatory Physiology model exposes future healthcare professionals to many physiological scenarios that their patients might experience, including the simulation of heart failure, anemia, and diabetes [18]. On the other hand, computer simulations are not simply an alternative tool to traditional instruction. Computer simulations can conceptually elucidate and quantitatively explore the fundamental science and mechanisms that comprise enormously complex biomedical systems, and this is making mathematical modeling and computer simulations an increasingly integral part of biological research [14]. Mathematical modeling and computer simulations, therefore, should be taught to undergraduate biology students

in an interactive manner, which allows students to directly implement and change the fundamental mathematics and computational implementation of these simulations. In this manner, educational computer simulations become not only a set of evolving pedagogical tools [14] but should reflect the inquisitive and investigative nature of computer simulations in biological research. Introducing computer programing to undergraduate biology students is not only important for establishing a future generation of computational biologists, however, but also improves students use of scientific logic and comfort with quantitative analysis [19], which are required skills for all future biologists [10].

Spreadsheets are important tools for biologists and an excellent tool for introducing biology students to mathematical modeling and computer simulations. Spreadsheets are a particularly useful teaching resource at the university level because essentially all classrooms in U.S. public schools (and classrooms in most developed countries) have access to computers and the internet [20, 21]. Futhermore, spreadsheets are increasingly being incorporated in to online environments (such as course management systems and ebooks). That said, biology students can still often struggle with using spreadsheet software and it is important that students recognize the difference between the science they are trying to elucidate and the computational mechanisms of the spreadsheet environment [19]. However, the widespread use of spreadsheets both in biological research, and in the workplace, make spreadsheets the obvious computational environment to introduce computer programing to undergraduate biology students. Spreadsheet models can also be very complex [22]. For example, Geyer recently implemented the Hodgkin-Huxley Model for action potentials in neurons using a spreadsheet model, allowing the students to explore this model and design their own experiments to further their understanding of neuron behaviour [23]. As another example, Langendorf and Strode recently implemented a simulation of an evolving population experiencing natural selection to introduce evolution in the classroom [24]. The inherent complexity can sometimes make the

behaviour of these systems difficult for both students and researchers to intuitively deduce, and the use of computer simulations can be invaluable to elucidating the nature of these systems. Furthermore, mechanisms within a computer model can be systematically (and sometimes unrealistically) turned off in a simulation to explore the strength and behaviour of different effects and interactions.

Here, we describe a computational biology course that introduces computer simulations to undergraduate biology students using a spreadsheet environment. The course both introduces the students to computer models, and also incorporates a research component that sees students implement computer models found within the scientific literature and adapt these computer models as part of an in-class scientific research project. Furthermore, a couple of examples of computer simulations that are implemented by our students as part of the course are detailed. The spreadsheet simulations described here are the Lotka-Volterra predator-prey model, a cellular automaton model of tumor growth, and a model of an infectious disease outbreak.

2 Computational Biology

Computational Biology is a course taught at our institution that uses spreadsheets to introduce mathematical modelling and computer simulations to our undergraduate biology students. Students investigate models that explore a variety of different biological systems. The prerequisites for this class is General Physics I (a calculus-based mechanics class) which our students will usually take in their first or second years at university. Students are, therefore, familiar with the concept of mathematical models of the physical world and will already have taken Calculus I, which covers differentiation and introduces integration. No prior programming experience is required before students take this class, which makes spreadsheets the perfect programming environment for this course. Students also work on a research project and learn how to present their results in a scientific

manner.

The first part of the course introduces students to spreadsheets. In particular, we describe the basics of entering data, entering formula and some of the functions available to spreadsheets. Furthermore, we discuss the inclusion of macros and the addition of a button within the spreadsheet that runs the macro; whilst spreadsheet syntax is generally universal, it should be noted that macros (or scripts) are handled differently for different spreadsheet software.

In terms of mathematical formula, our students have not been introduced to finite difference calculations and numerical integration before taking this course. Therefore, the course covers the derivation of forward, backward and central difference approximations and the trapezoidal rule. It is important for students to understand the role that discretization has when numerically simulating a continuum mathematical model. In particular, the role of the temporal and spatial discretizations in the numerical stability of the computer simulation is an area that should be emphasized.

As an example we first consider diffusion in one-dimension. The diffusion equation is given by

$$\frac{\partial \phi}{\partial t} = \frac{\partial}{\partial x} D \frac{\partial \phi}{\partial x} \tag{1}$$

where ϕ is the concentration, t is time, D is the diffusion coefficient, and x is position. The discrete form of the equation that is implemented into the spreadsheet is of the form

$$\phi_i^{t+1} = \phi_i^t + D \frac{\Delta t}{(\Delta x)^2} \left[\phi_{i+1}^t + \phi_{i-1}^t - 2\phi_i^t \right] \tag{2}$$

where the superscript represents the discrete time, and the subscript represents the discrete location. This gives students a way to explore numerical stability and understand how the discretization in time and space, Δt and Δx respectively, influence numerical stability.

Another simple example that we initially consider is the logistic growth model, which captures the growth of a population that is constrained by resources or its environment. The rate equation is of the form

$$\frac{dN}{dt} = rN\left(\frac{K-N}{K}\right) \tag{3}$$

where N is the population, r is the growth rate, K is the carrying capacity and t is time. The discrete form of this equation is

$$N^{t+1} = N^t + \Delta t r N^t \left(K - N^t\right)/K \tag{4}$$

where the superscript represents the discrete time. This is a good example of growth rate that students may have seen in previous courses (without really looking too closely at the mathematics, and certainly without numerically implementing this model in a spreadsheet).

The course also introduces students to research methods and the process of writing scientific papers. Students would typically take this class prior to starting their senior thesis research project. The course then progresses by alternating each week between students working on their in-class research project and students implementing a model that is chosen by the instructor. The following sections give examples of some of the models that students implement in this course.

3 Lotka-Volterra predator-prey model

The Lotka-Volterra Predator-Prey System is usually captured with the following nonlinear differential equations that can result in a continual cycle of growth and decline [25]. The rate of change of the number of prey is given by

$$\frac{dx}{dt} = \alpha x - \beta xy \tag{5}$$

where x is the number of prey, y is the number of predators, α is the prey growth rate and β is the rate of predation. The rate of change of the number of predators is given by

$$\frac{dy}{dt} = \delta xy - \gamma y \tag{6}$$

where δ is the predator growth rate and γ is the predator loss rate. These simple coupled equations can lead to quite interesting behaviour.

These equations can be discretized and implemented in a spreadsheet environment by the students to explore the complex interactions between predators and prey in this model. The discretized equations are simply

$$x^{t+1} = x^t + \Delta t \left(\alpha x^t - \beta x^t y^t\right) \tag{7}$$

and

$$y^{t+1} = y^t + \Delta t \left(\delta x^t y^t - \gamma y^t\right) \tag{8}$$

where Δt is the time step and superscripts represent discrete time. An implementation of this model is depicted in Figure 1. This is a wonderful example of the dynamics of a model which undergoes cyclic behaviour. For undergraduate biology students, whose only experience of calculus is a course that covers very basic differentiation (always with respect to one variable) and a brief introduction to integration, implementing these coupled equations and observing the resultant dynamic behaviour, can be quite thought-provoking. Once students have implemented this model they can try to see if they can capture the dynamics of a real system; real systems invariably exhibit greater complexity than the models we use to mimic their behaviour and this can result in a good area of discussion. For example, some of my students had already taken a zoology course with a colleague whose research interests is in amphibians, and this resulted in a classroom discussion about the applicability of

Lotka-Volterra Predator-Prey System	
Prey Growth Rate	3
Rate of Predation	1
Predator Growth Rate	0.1
Predator Loss Rate	1
Prey	Predators
10	10
9.3	10
8.649	9.93
8.0496243	9.7958457
7.502584253	9.604789906
7.007054326	9.36491837
6.561061038	9.084631451
6.161844654	8.77221652
5.806169639	8.435525223
5.490573824	8.081753605
5.211556391	7.717312893
4.965710969	7.347773717
4.749813093	6.97786555
4.560871914	6.611514567
4.396155361	6.251905821

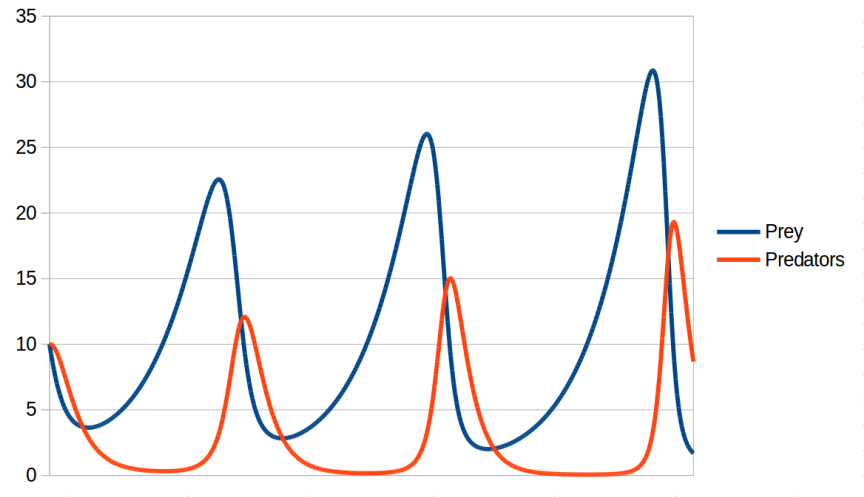


Figure 1: Example of the spreadsheet implementation of the Lotka-Volterra predator prey model. The system is cyclic. As the predators eat the prey their population increases, but the number of prey decreases. This is unsustainable and the number of predators will decrease once the number of prey is sufficiently low. However, as the number of predators decreases the number of prey is allowed to rebound, and with it the number of predators, and the cycle reiterates once more.

this model to wood frogs. Wood frogs are found in Western Pennsylvania and have an interesting ability to tolerate freezing temperatures. This allows the frogs to stay nearer the surface during winter and emerge sooner in the spring and capture the prey first, before other frogs that have to bury themselves further underground during the winter. The class discussion centered around how we can mimic the effects of temperature in the model. The students were interested in the idea that the properties of the model could be time-dependent and the "activity" of the freeze-tolerant frogs and various insects could be slowly turned on to mimic the beginning of spring.

4 Cellular automaton model of tumor growth

This two-dimensional model of cancer growth is adopted from Poleszczuk and Enderling [26] and can easily be implemented, on a smaller scale, within a spreadsheet model. Within this spreadsheet model, each cell of the spreadsheet represents an individual cancer cell that occupies a spatial area

of $(10\,\mu\text{m})^2$ on a two-dimensional regular square lattice.

Each cancer cell is characterized by its remaining proliferation potential (how many times left that a cell can divide), ρ, and its probability of spontaneous death α. The model assumes a heterogeneous tumor population consisting of cancer stem cells and non-stem cancer cells. Cancer stem cells are assumed to be immortal and have unlimited proliferation potential. In other words, their remaining proliferation potential is infinite, and their probability of spontaneous death is zero. Non-stem cancer cells, on the other hand, can only divide a limited number of times, ρ_{max}, before cell death. Each cell type can divide symmetrically to produce two daughter cells with parental phenotype. In other words, a stem cell would proliferate two stem cells, but a non-stem cell would proliferate two non-stem cells (with a remaining proliferation potential reduced by 1). However, a cancer stem cell can also undergo asymmetric division and create a cancer stem cell and a non-stem cancer cell with $\rho = \rho_{max}$ (and the remaining proliferation potential of these non-stem cells would decrease with each subsequent non-stem cell division like any other non-stem cell). The probability of asymmetric division is $1 - p_{symm}$ (where p_{symm} is the probability of symmetric cancer stem cell division). Cells need adjacent space for proliferation, and cells that are completely surrounded by other cells (the surrounding eight spreadsheet cells on our two-dimensional lattice) become quiescent. Otherwise, cells can randomly proliferate into vacant adjacent space. Cells can undergo spontaneous death, independent of the available space, with rate of α. Cells that die are instantaneously removed from the system, and this space is then considered empty. The temporal discretization in the model is $\Delta t = 1/24$ day (i.e., 1 hour), and proliferation probabilities are scaled to this simulation time step.

A layout of the simulation is depicted in Figure 2. The simulation (which consists of 20×20 cells in the spreadsheet) is replicated 14 times through the spreadsheet. The first instance is shown as the left-most square in Figure 2 and contains the current state of the system. This is comprised

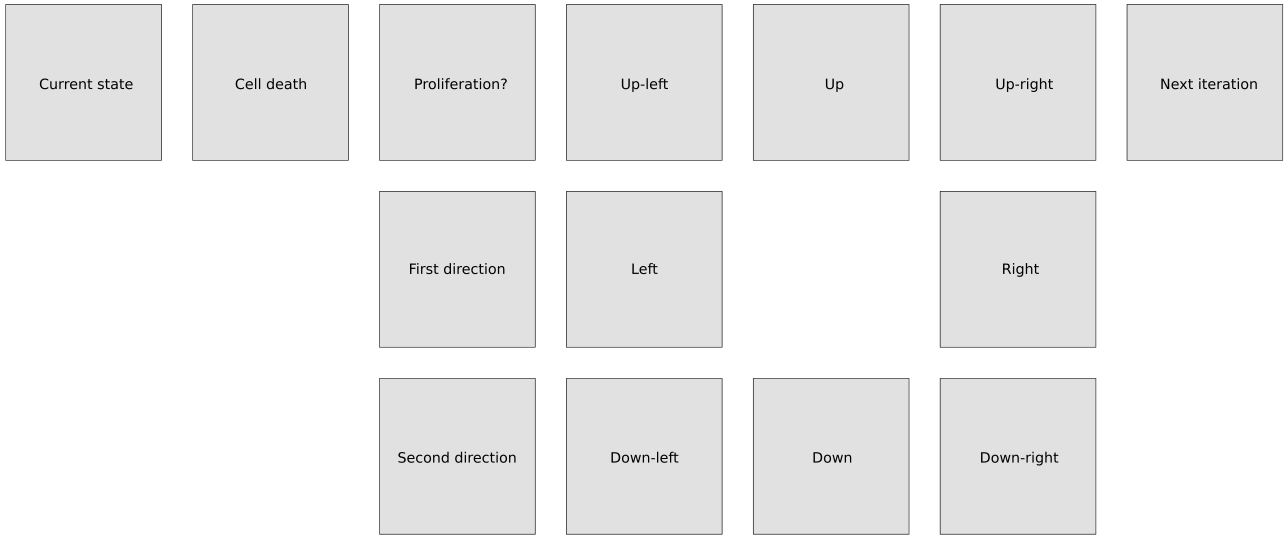


Figure 2: Example of the spreadsheet implementation of a two-dimensional cellular automata model of cancer growth. This is how the different stages of the variables during a single iteration are stored across the spreadsheet. Starting on the left is the current state of the system, and moving across to the right we end at the next iterative state.

of a number which represents the current remaining proliferation potential (or -1 if the spreadsheet cell contains a cancer stem cell). Immediately to the right of this is the next 20×20 grid that contains either 0 if the cell is going to spontaneously die or maintains the remaining proliferation potential if the cell survives onto the proliferation stage of the simulations iteration. For any cells which are non-stem cancer cells we check to see if a random number is less than the probability of the cell dying (in other words, if the random number is less than α then the cell is removed).

Next, we have to capture the proliferation of the cancer cells. If a random number is less than the probability of proliferation and the space is occupied than the cell is identified to proliferate (the value within the spreadsheet cell is set to 1). Below this is another two 20×20 grids which, if the cell is identified to proliferate, is assigned a value of between 1 and 8 that represent directions that the cancer cell will attempt to proliferate in.

In figure 2 there are now a series of eight 20×20 grids arranged in the directions that the proliferation can be in. We go from one direction to the next and if the cell is empty, and the cell in

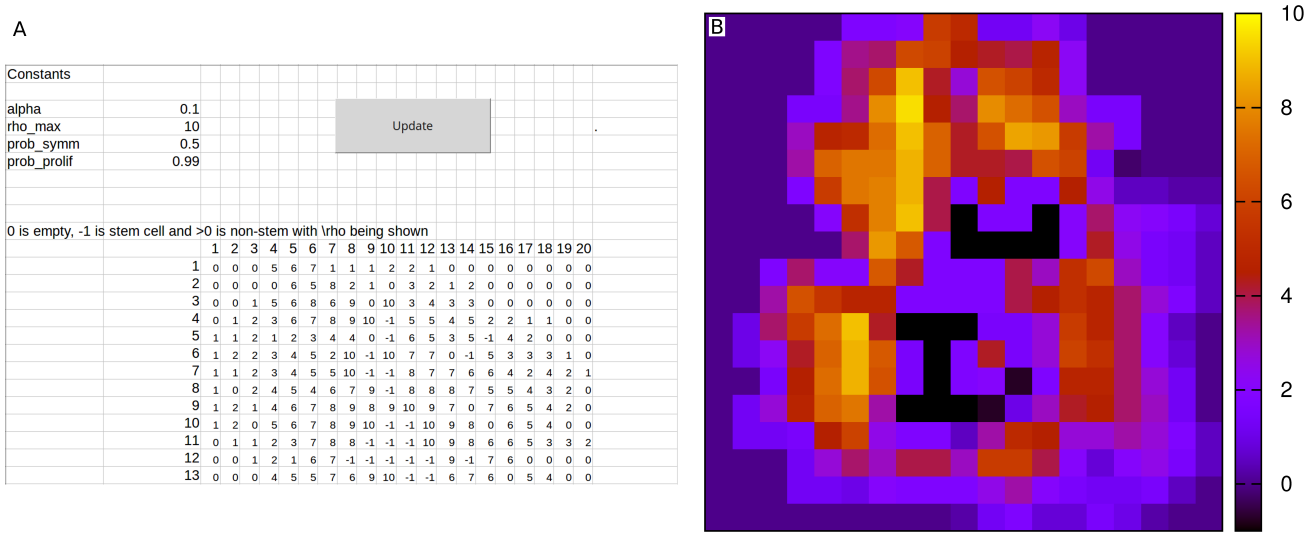


Figure 3: Example of the spreadsheet implementation of a two-dimensional cellular automata model of cancer growth. a) A screenshot of the spreadsheet, showing the variables at the top and the current state of the simulation below, and b) a contour plot of the remaining proliferation potential (or -1 if the spreadsheet cell contains a cancer stem cell).

the opposite direction is proliferating in this direction then it will contain a new cell. The identity of the new cell is randomly determined to be either symmetrically proliferating or asymmetrically proliferating (creating both stem and non-stem cells) based on the probability p_{symm}. The new remaining proliferation potential is determined (or the spreadsheet cell contains -1 if the cancer stem cell proliferates another cancer stem cell). In other words, if the cell is occupied by a stem cancer cell then the new cell can either by a stem cell (equal to -1) or a non-stem cell with a proliferation potential set to the maximum value, but if it is occupied by a non-stem cell then the proliferation potential of the newly created non-stem cell is reduced by 1. Note the probability of the stem cancer cell proliferating either a stem cell or a non-stem cell depends on the probability of symmetric proliferation This is calculated for each direction in turn and then copied to the last 20×20 grid, on the right of Figure 2, which represents the next iteration of the simulation.

Now to complete an iteration we just have to copy the next iteration grid over to the original grid. We can create a macro that will simply copy the contents and paste special (the values but

not the formulas) to the original cells, and insert a button that when pressed will run the macro to update the system. Figure 3a shows a snapshot of the spreadsheet simulation. The constants are stored at the top of the spreadsheet and the current state of the system is below it (the top left 20×20 grid from Fig. 2). At the top of the screenshot is the button that will run a macro that will copy the next iteration values back to the current iteration values. Once these values are copied the spreadsheet will automatically recalculate all of the values for the next iteration. In other words, every time the button is pressed the simulation progresses through one iteration (representing an hour of time). Figure 2b shows the remaining proliferation potential (or -1 if the spreadsheet cell contains a cancer stem cell) after a number of iterations. For the values considered here there would appear to be a greater concentration of stem cancer cells in the center of the tumor.

During the class discussions that have followed students seem skeptical and believe that the model might not necessarily represent a real tumor. In particular, the two-dimensional array of numbers in the spreadsheet can be difficult for students to visualize as an actual representation of a tumor. Discussions in this class can sometimes be a little uncomfortable as there will be students who have lost family or friends to cancer. I generally steer the conversation to cancer treatments. For example, using nanoparticles to deliver chemotherapeutic agents [27], immunotherapy [28] or molecularly targeted therapy [29]. The model predicts that the stem cells are generally contained in the center of the tumor, and this has lead the discussion on to the accessibility of different parts of the tumor to nutrients (and the administered drugs) and the emergence of a complex vasculature in tumors. Interestingly this can result in an enhanced permeability and retention (EPR) effect which sees small particles tending to accumulate in tumor tissue to a much greater extent than they do in normal tissues. Furthermore, there is a strong debate on the nature of the emergence of drug resistance during chemotherapy that is worth mentioning. The debate surrounds whether drug resistance is initially present in the tumor to some small extent, or if the drug resistance emerges

while the tumor is exposed to treatments than can increase mutation rates. In the latter case, does having stem cells in the center of a tumor surrounded by non-stem cancer cells that are being exposed to chemotherapeutic agents result in a greater probability of drug resistance emerging?

5 Model of an infection outbreak

The models that represent the outbreak of an infection are well-established and this computer simulation is used to introduce these models to the students. We choose an infection that the students are familiar with for this example, and simulate the spread of a zombie apocalypse [30]. This lecture coincides with the week of Halloween and is a fun way for the students to explore this class of model. Furthermore, students often get excited about which zombie movie or television show is their favourite, and attempt to vary the parameters to best replicate the effects of their chosen strain of zombie outbreak.

The differential equations for this model are

$$\frac{dS}{dt} = \Pi - \beta SZ - \delta S$$

$$\frac{dI}{dt} = \beta SZ - \rho I - \delta I$$

$$\frac{dZ}{dt} = \rho I - \alpha SZ$$

and

$$\frac{dR}{dt} = \delta S + \delta I + \alpha SZ$$

where S is the number of susceptible people, I is the number of people infected, Z is the number of zombies and R is the number of people permanently removed from the system (dead, but not

undead). Π is the birth rate of new people, β is the transmission factor (how people become infected), δ is the rate at which people die from non-zombie related causes, ρ is the rate at which the infected become actual flesh-eating zombies, and α is the rate at which humans kill the zombies.

We will solve these equations using the usual finite difference equations (Euler method) and choose our time step, Δt, to be suitably small. The discretized equations, that are suitable for incorporating into the spreadsheet, are

$$S_{t+1} = S_t + \Delta t \left[\Pi - \beta S_t Z_t - \delta S_t\right]$$

$$I_{t+1} = I_t + \Delta t \left[\beta S_t Z_t - \rho I_t - \delta I_t\right]$$

$$Z_{t+1} = Z_t + \Delta t \left[\rho I_t - \alpha S_t Z_t\right]$$

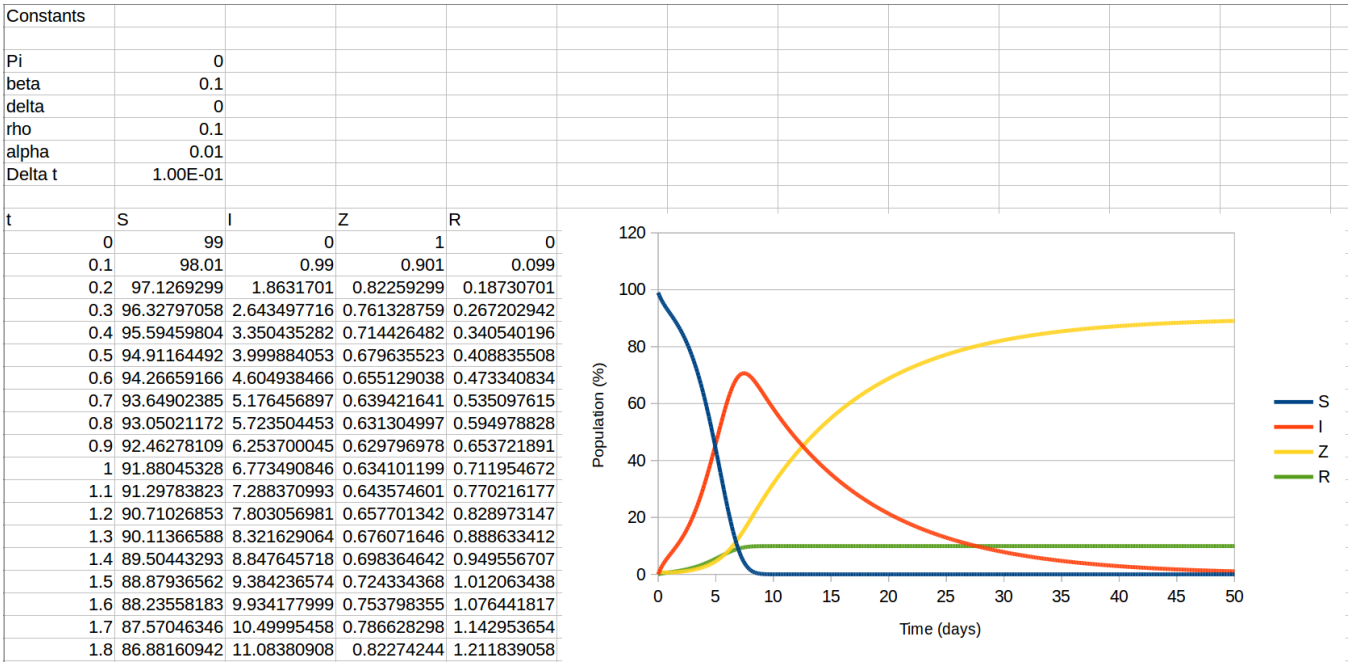
and

$$R_{t+1} = R_t + \Delta t \left[\delta S_t + \delta I_t + \alpha S_t Z_t\right]$$

where the subscripts represents the time. A screenshot of the spreadsheet simulation is included as Figure 4. The various constants are placed at the top of the spreadsheet and directly below this are 5 columns. The first column is time, with each subsequent cell just being updated by the value of Δt. The second column is the number of susceptible people and this is increased by the birthrate (which is relatively small), significantly decreased by the transmission of the zombie "disease" and slightly decreased by the rate at which people die from non-zombie related causes. The third column is the number of people who are infected with the zombie disease (bitten by a zombie). This is increased by the same rate of transmission of the zombie "disease" that significantly decreased the number of susceptible people, but is also decreased by the rate at which the infected become zombies and (similar to the susceptible population) slightly decreased by the rate at which people

die from non-zombie related causes. The fourth column is the number of zombies. This is increased as the infected turn into zombies and is only decreased by the rate at which susceptible people kill the zombies (although students who reference the movie 28 days later argue that over large times the zombies will die off on their own). The final column is the number of people removed from the system. This is increased by the death of susceptible and infected people, or the death of zombies at the hand of normal humans; note that in the current model this saturates at α/β, as with the variables considered here the death of zombies is much more prominent than the regular death rate (although we might expect this to increase during the apocalypse). The zombies quite quickly take over the world in the current model.

This provides a wonderful way for students to interact with this model and understand the relationship between the terms in the equations and the overall behaviour of the system. This model also allows students the opportunity to add additional terms to the equations (or even populations) to make the model more accurately capture the behaviours found in their favourite zombie movies. As an example I generally start the discussion by comparing Night of the Living Dead slow-moving zombies to the 28 Days Later fast-moving zombies. In the case of fast-moving zombies we might expect the probability of zombies infecting humans to be much higher and the probability of humans killing zombies to be much smaller. Furthermore, the George A. Romero zombies can infect susceptible people and the infection leads to zombification over several hours, while Alex Garland imagined the zombie transformation to be almost instantaneous. This can be included by drastically increasing the rate that infected are transformed to zombies. The most interesting discussions, in my opinion, relate to areas that are not initially part of the model. For example, in one class a couple of students modified the model to account for susceptible people killing infected people. I think this is a really interesting development. Not just taking a model and changing the parameters, but exhibiting the confidence and creativity to modify the model to



Constants				
Pi	0			
beta	0.1			
delta	0			
rho	0.1			
alpha	0.01			
Delta t	1.00E-01			
t	S	I	Z	R
0	99	0	1	0
0.1	98.01	0.99	0.901	0.099
0.2	97.1269299	1.8631701	0.82259299	0.18730701
0.3	96.32797058	2.643497716	0.761328759	0.267202942
0.4	95.59459804	3.350435282	0.714426482	0.340540196
0.5	94.91164492	3.999884053	0.679635523	0.408835508
0.6	94.26659166	4.604938466	0.655129038	0.473340834
0.7	93.64902385	5.176456897	0.639421641	0.535097615
0.8	93.05021172	5.723504453	0.631304997	0.594978828
0.9	92.46278109	6.253700045	0.629796978	0.653721891
1	91.88045328	6.773490846	0.634101199	0.711954672
1.1	91.29783823	7.288370993	0.643574601	0.770216177
1.2	90.71026853	7.803056981	0.657701342	0.828973147
1.3	90.11366588	8.321629064	0.676071646	0.888633412
1.4	89.50443293	8.847645718	0.698364642	0.949556707
1.5	88.87936562	9.384236574	0.724334368	1.012063438
1.6	88.23558183	9.934177999	0.753798355	1.076441817
1.7	87.57046346	10.49995458	0.786628298	1.142953654
1.8	86.88160942	11.08380908	0.82274244	1.211839058

Figure 4: Example of the spreadsheet implementation of a model of an infection outbreak. The infection is chosen to be a zombie apocalypse and the number of susceptible, infected, zombified, and dead people are calculated and plotted as a function of time.

capture the behaviour they want to mimic.

6 Summary and conclusions

Here we have described a course which we offer that introduces our undergraduate biology students to computational thinking, mathematical modeling, computer simulations and the power of spreadsheets in implementing simple models. The course stresses the role of computer simulations in scientific research, and the benefits of computer literacy and quantitative skills throughout biological research. Three examples of models implemented in a spreadsheet environment by our students have been detailed: the Lotka-Volterra predator-prey model, a cellular automaton model of tumor growth, and a model of an infectious disease outbreak. Other computer simulations also currently explored within this course include the dynamic instability of microtubules [31], the treatment of drug-resistant strains of tumor cells [32], a tissue heat transfer model of scalds and burns [33], and

a model of dry-eye syndrome [34]. However, the possibilities are endless and the choice of models to explore within the class is largely based on the interests of the instructor and students.

We have found that the incorporation of computer simulations within undergraduate biology is important for improving the quantitative skills of our students, and their ability to use spreadsheets to analysis experimental results. This is perhaps not too surprising after an entire semester course where the students implement computer models in a spreadsheet environment. Furthermore, a number of students have subsequently gone on and incorporated spreadsheet models within their senior thesis research projects and honors research projects (outside of this course). In particular, in the last two sections offered (which contained on average 15 students each) 6 biology students have continued using computer models and simulations in their future research projects; something that would have been unheard of before this class was offered. The long term impact of this course is, therefore, expected to be significant in the research abilities and possible research direction of our graduating students. Certainly we hope that the observation that biologists are uncomfortable with articles that contain mathematical models and computer simulations would not apply as much to our students after taking this course. As computer simulations and computational thinking become increasingly prevalent in K-12 education [35] we might expect that undergraduate biology students will become increasingly comfortable with integrating quantitative skills into their biology research. Socioeconomic factors will see this occur in some demographics sooner than others, and including a comprehensive introduction to computer simulations within an undergraduate biology sequence is expected to help diversify the make up of future computational biologists. Regardless of a student's background, however, I have observed that all students that have taken this course have developed an ability to implement and manipulate computer models within a spreadsheet environment, and I believe that their general spreadsheet skills have improved significantly.

References

[1] Cynthia DAngelo, Daisy Rutstein, Christopher Harris, R Bernard, E Borokhovski, and G Haertel. Simulations for STEM learning: Systematic review and meta-analysis. *Menlo Park: SRI International*, 2014.

[2] Sapna Cheryan, Sianna A Ziegler, Amanda K Montoya, and Lily Jiang. Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, 143(1):1, 2017.

[3] Stephen J Farenga and Beverly A Joyce. Intentions of young students to enroll in science courses in the future: An examination of gender differences. *Science Education*, 83(1):55–75, 1999.

[4] Kevin S Bonham and Melanie I Stefan. Women are underrepresented in computational biology: An analysis of the scholarly literature in biology, computer science and computational biology. *PLoS computational biology*, 13(10):e1005134, 2017.

[5] Suzanne K Damarin. Technology and multicultural education: The question of convergence. *Theory into Practice*, 37(1):11–19, 1998.

[6] Christopher Ball, Kuo-Ting Huang, RV Rikard, and Shelia R Cotten. The emotional costs of computers: an expectancy-value theory analysis of predominantly low-socioeconomic status minority students STEM attitudes. *Information, Communication & Society*, pages 1–24, 2017.

[7] National Research Council et al. *A new biology for the 21st century*. National Academies Press, 2009.

[8] Terry Woodin, V Celeste Carter, and Linnea Fletcher. Vision and change in biology undergraduate education, a call for actioninitial responses. *CBE-Life Sciences Education*, 9(2):71–73, 2010.

[9] Louis J Gross. Education for a biocomplex future. *Science*, 288(5467):807–807, 2000.

[10] Jason Feser, Helen Vasaly, and Jose Herrera. On the edge of mathematics and biology integration: improving quantitative skills in undergraduate biology education. *CBE-Life Sciences Education*, 12(2):124–128, 2013.

[11] Tim W Fawcett and Andrew D Higginson. Heavy use of equations impedes communication among biologists. *Proceedings of the National Academy of Sciences*, 109(29):11735–11739, 2012.

[12] William R Robinson. The inquiry wheel, an alternative to the scientific method. a view of the science education research literature, 2004. *Journal of Chemical Education*, 81(6):791–792, 2004.

[13] Dara B Duncan, Alexandra Lubman, and Sally G Hoskins. Introductory biology textbooks under-represent scientific process. *Journal of Microbiology & Biology Education: JMBE*, 12(2):143, 2011.

[14] Tomáš Helikar, Christine E Cutucache, Lauren M Dahlquist, Tyler A Herek, Joshua J Larson, and Jim A Rogers. Integrating interactive computational modeling in biology curricula. *PLoS computational biology*, 11(3):e1004131, 2015.

[15] M Campbell, LJ Heyer, and CJ Paradise. Integrating concepts in biology, 2014.

[16] Nicola J Gibbons, Chris Evans, Annette Payne, Kavita Shah, and Darren K Griffin. Computer simulations improve university instructional laboratories. *Cell Biology Education*, 3(4):263–269, 2004.

[17] Joseph P Akpan. Issues associated with inserting computer simulations into biology instruction: a review of the literature. *Electronic Journal of Science Education*, 5(3), 2001.

[18] Sean R Abram, Benjamin L Hodnett, Richard L Summers, Thomas G Coleman, and Robert L Hester. Quantitative circulatory physiology: an integrative mathematical model of human physiology for medical education. *Advances in Physiology Education*, 31(2):202–210, 2007.

[19] Karen L Carleton, Carly H Rietschel, and Gili Marbach-Ad. Group active engagements using quantitative modeling of physiology concepts in large-enrollment biology classes. *Journal of microbiology & biology education*, 17(3):487, 2016.

[20] Thomas D Snyder and Cristobal de Brey and Sally A Dillow. Digest of Education Statistics 2014, NCES 2016-006. *National Center for Education Statistics*, 1016.

[21] European Schoolnet. Survey of schools: ICT in education. Benchmarking access, use and attitudes to technology in European schools. *Liége, Belgium: European Union. doi*, 10: 94499, 2013.

[22] Gavin Buxton. *Alternative Energy Technologies: An Introduction with Computer Simulations*. CRC press, 2017.

[23] Florian Henning Geyer. A spreadsheet implementation of the Hodgkin-Huxley model for action potentials in neurons. *Spreadsheets in Education (eJSiE)*, 10(1):1, 2017.

[24] Ryan E Langendorf and Paul K Strode. Using spreadsheets to simulate an evolving population. *The American Biology Teacher*, 79(8):635–643, 2017.

[25] Alan A Berryman. The orgins and evolution of predator-prey theory. *Ecology*, 73(5):1530–1535, 1992.

[26] Jan Poleszczuk and Heiko Enderling. A high-performance cellular automaton model of tumor growth with dynamically growing domains. *Applied mathematics*, 5(1):144, 2014.

[27] Gavin A Buxton. Simulating the co-encapsulation of drugs in a smart core-shell-shell polymer nanoparticle. *The European Physical Journal E*, 37(3): 14, 2014.

[28] Steven A Rosenberg, James C Yang, and Nicholas P Restifo. Cancer immunotherapy: moving beyond current vaccines. *Nature medicine*, 10(9): 909, 2004.

[29] Sith Sathornsumetee, David A Reardon, Annick Desjardins, Jennifer A Quinn, James J Vredenburgh, and Jeremy N Rich. Molecularly targeted therapy for malignant glioma. *Cancer*, 110 (1): 13–24, 2007.

[30] Philip Munz, Ioan Hudea, Joe Imad, and Robert J Smith?. When zombies attack!: mathematical modelling of an outbreak of zombie infection. *Infectious disease modelling research progress*, 4:133–150, 2009.

[31] Gavin A Buxton, Sandra L Siedlak, George Perry, and Mark A Smith. Mathematical modeling of microtubule dynamics: insights into physiology and disease. *Progress in neurobiology*, 92(4):478–483, 2010.

[32] Mitra Shojania Feizabadi and Tarynn M Witten. Modeling drug resistance in a conjoint normal-tumor setting. *Theoretical Biology and Medical Modelling*, 12(1):3, 2015.

[33] Cameron Loller, Gavin A Buxton, and Tony L Kerzmann. Hot soup! correlating the severity of liquid scald burns to fluid and biomedical properties. *Burns*, 42(3):589–597, 2016.

[34] Daniel Anderson, Katlyn Winter, and Richard Braun. A model for wetting and evaporation of a post-blink precorneal tear film. In *APS Division of Fluid Dynamics Meeting Abstracts*, 2009.

[35] Donna L Cellante, Sushma Mishra, Benjamin R Campbell, Mary A Hansen, and Gavin A Buxton. The perceptions of middle school teachers about the integration of STEM + C: A focused-group approach. *Issues in Information Systems*, 18(1), 2017.