

# Project Reports

## Mathematical Modeling with Dr. Alber

Daniel Collister and Charlie Holderman

Winter 2019

# Project 1: Cellular Automata

## Introduction

This is a model of a 1-D cellular automata consisting of two interacting populations with periodic boundary conditions. These periodic boundary conditions mean that the domain on which this simulations is taking place is topologically equivalent to the boundary of a circle. This model consists of two populations (of cells), A and B. B cells are able to reproduce and move at unit speed. A cells are not able to reproduce, but are able to recruit a B cells to switch to A cells. This happens when two A cells surround a B cell. A cells are able to move at twice the speed of B cells. Both populations die at the same rate.

To handle the spatial component of this model, we make the assumption that cells are tightly packed in together, so that there are no empty spaces between cells. Our model is essentially a lattice-based model, but the number of lattice sights changes as the total cell population increases or decreases. Thus, a cells position is not absolute, but relative to that of other nearby cells.

At each time step all cells move randomly, B cells up to one space right or left (or no movement), and A cells up to two spaces right or left (or no movement). The new (relative) positions are then calculated. Any B cells whose immediate neighbors are A cells switch to become A cells. Next, for every pair of adjacent B cells, there is a random number generated. If this number is less than the birth rate,  $\beta$ , then a new B cell is added in between the previously adjacent cells. In other words,  $P(\text{birth}) = \beta$ . Finally, cell death is checked (randomly) for all cells in a similar way where  $P(\text{death}) = \mu$ .

## Sensitivity Analysis

Simulations were run to test the effect of the parameters dictating birth and death rates,  $\beta$  and  $\mu$ , respectively. The range of each parameter was  $\beta, \mu \in [0, 1]$ . A total of 10,000 parameter pairs were tested (each parameter taking values from 0 to 1 in increments of 0.01).

For a fixed value of the parameter pair  $(\beta, \mu)$ , the cellular automata model was run in the following manner. Let  $N_0$  denote the size of the population at time  $t = 0$ . (Note that the size of the entire population can increase or decrease as new cells are born or die off). For our simulations,  $N_0 = 50$ . The initial configuration of the population is chosen randomly. For  $i = 1, \dots, N_0$ , the  $i$ -th cell has equal probability of being A or B, i.e.  $P_i(A) = P_i(B) = 0.5$ .

The model was then allowed to run according to the above description for a maximum of 500 time steps or until one of the cell populations died out completely. The final ratio of A to B cells was recorded as follows. If we let  $c_i$  denote the  $i$ -th cell, and  $N$  denote the total population size then we can define the population of A cells at the end of the simulation by

$$A_\infty = \{c_i = a \mid i = 1, \dots, N\}$$

and similarly for  $B_\infty$ , then

$$R_\infty^A = \frac{|A|}{|A| + |B|}; \quad R_\infty^B = \frac{|B|}{|A| + |B|}.$$

The simulation was run a total of 50 times, each with a different random starting configuration.  $R_\infty^A$  was recorded and the average and variance across the 50 simulations was calculated and saved for the parameter pair  $(\beta, \mu)$ .

## Results

The goal of the above process was to find a range of parameter values for which a non-trivial equilibrium state was reached, i.e. neither the A nor B population went extinct. Figure 1 shows the result of this process for a range of parameter values  $(\beta, \mu)$ . We can see three distinct regions, two of which are interesting. Note that blue corresponds to B cells survival and yellow to A cell survival.

First, the uninteresting region is shaded blue (B cells survive) on the right side of the graph. This region is one with a very high death rate ( $\mu > 0.5$ ) so eventually both populations go extinct. In this region, A cells tend to go extinct first, presumably because they cannot reproduce themselves, whereas B cells are able to reproduce, thus coloring the region blue.

The two interesting regions are the roughly triangular regions on the left. The first of these areas (yellow region, lower left) tends to fairly robustly favor survival of A cells and the extinction of B cells. This corresponds to low birth rate and low death rate. The other region corresponds to low death rate, high birth rate and favors B cell survival.

Our best chance at finding a non-trivial equilibrium state would be found along the boundary where these two regions meet. In fact, at first glance, we see that along this boundary  $R_\infty^A \approx 0.6$ , which could indicate a stable equilibrium of A and B cells together. However, we recall that this average is taken across 50 simulations and look at the variance of  $R_\infty^A$  for these simulations in Figure 7 to see that this boundary region is also the region of highest variance. The conclusion to be drawn from this is that we do not, in fact, have a stable equilibrium where A and B cells exist alongside each other, but rather an unstable region where for any given simulation, we have an equal probability of tending towards an A extinction or a B extinction. Thus, we seem to have found conclusive evidence that we cannot find a region with a non-homogenous equilibrium state.

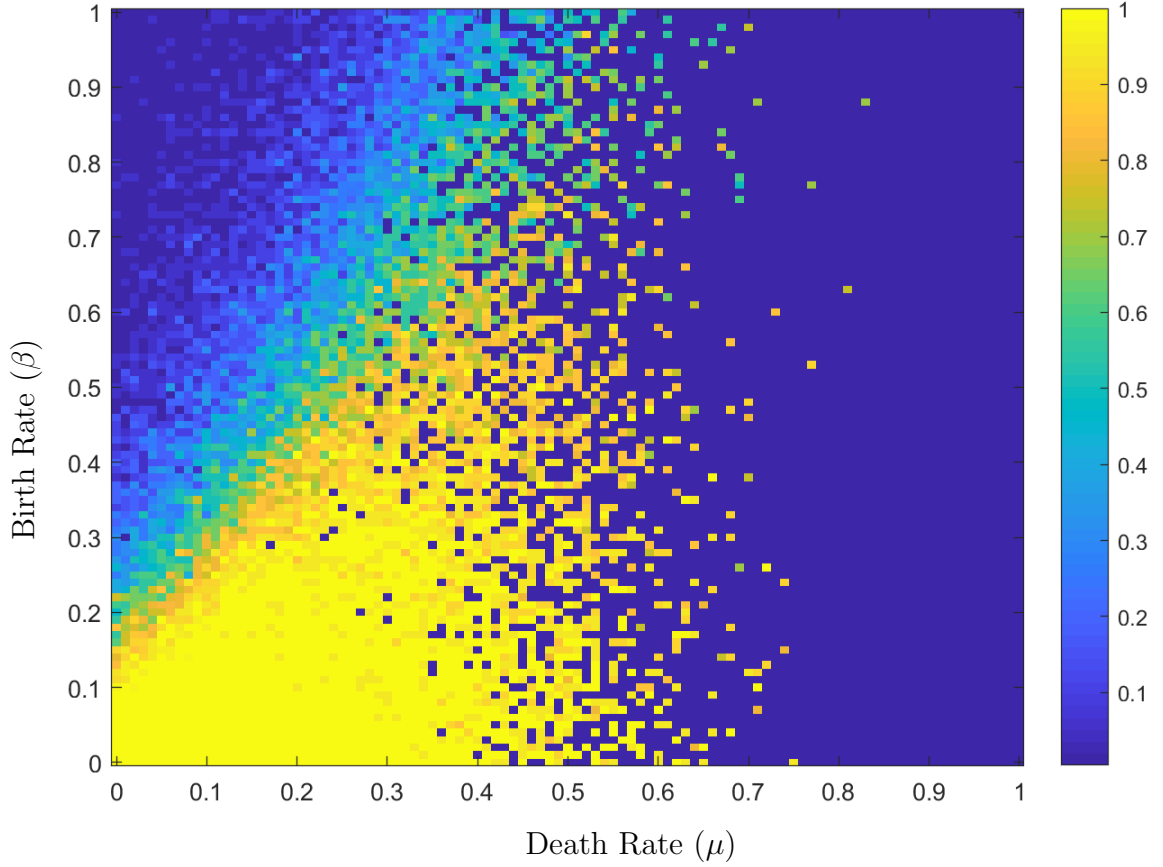


Figure 1: Average value of  $R_{\infty}^A$  for 50 simulations per parameter pair.

# Project 2: Ising Model

## Introduction

This project was a 2-D Ising model run on a  $60 \times 60$  grid with periodic boundary conditions. Each cell in the grid could take the values  $\pm 1$  representing ON or OFF, respectively, i.e. if  $c_{i,j}$  denotes the cell at position  $(i,j)$ , then  $c_{i,j} \in \{1, -1\}$ . The model is run according to a Monte Carlo Method. That is, at each step a cell is chosen at random and “flipped” from ON to OFF, or vice versa. If this change results in a more favorable energy configuration, then the change is accepted, if not, it may still be accepted with some prescribed probability. More specifically, for a given energy function  $E[\{s_i\}]$ , define

$$\Delta E := E_{\text{new}} - E_{\text{old}}.$$

Then the probability of a change being accepted is given by

$$P(\text{flip}) = \min \{1, e^{-\Delta E}\}.$$

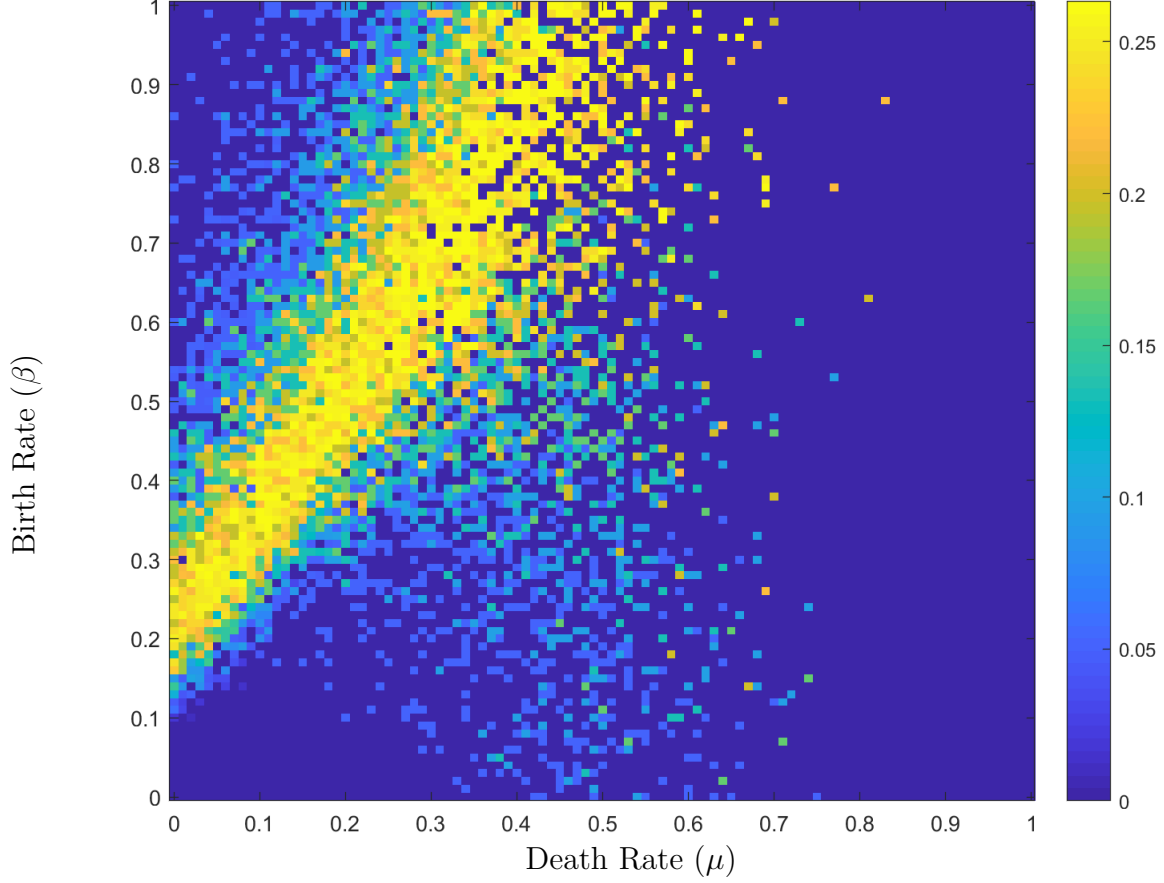


Figure 2: Variance of  $R_\infty^A$  for 50 simulations per parameter pair.

## Methods

The backbone of this model is the energy function  $E[\{s_i\}]$ , so we discuss it here. This function is defined as follows:

$$E[\{s_i\}] = -h \sum_i s_i - \frac{1}{2} \sum_{i,j} J(r_{i,j}) s_i s_j.$$

Here  $h$  is a parameter that represents towards ON or OFF depending on its sign,  $r_{i,j}$  is the distance between the  $i$ -th and  $j$ -th cells, and  $J(r_{i,j})$  is a function dictating the interaction between cells that depends on this distance. In our model, we took  $J$  to be a step function, where for some cutoff radii  $R_1 < R_2$ ,  $J$  is constant within each radius and zero outside, in other words:

$$J(r_{i,j}) = \begin{cases} J_1, & 0 < r_{i,j} \leq R_1 \\ J_2, & R_1 < r_{i,j} \leq R_2 \\ 0, & R_2 < r_{i,j}. \end{cases}$$

The simulation begins with a randomly chosen initial configuration. The simulation continues to run until there is very little energy change for a large number of iterations, currently set to 300 iterations. Note that the energy change will never be equal to zero, since unfavorable energy changes are accepted probabilistically. There is also a maximum run time of 100,000 iterations. These conditions together are almost always sufficient to view the resulting patterns.

## Results

To view the results of these simulations, there are two options. First, to watch the simulation with a randomly chosen set of parameter values, one simply needs to run the file “testRandomParameter.m.” Running this file chooses the parameters randomly from a prescribed range, so successive runs of this file should produce vastly different results. The second option is to run the file “runSavedParameter.m.” This file contains saved parameter values that result in qualitatively different (and interesting!) patterns. To view the different saved patterns, one simply needs to comment/uncomment lines in the first section so that only one line of ‘params’ is uncommented.

Below are some examples of the wide range of patterns generated, along with the parameter values used.

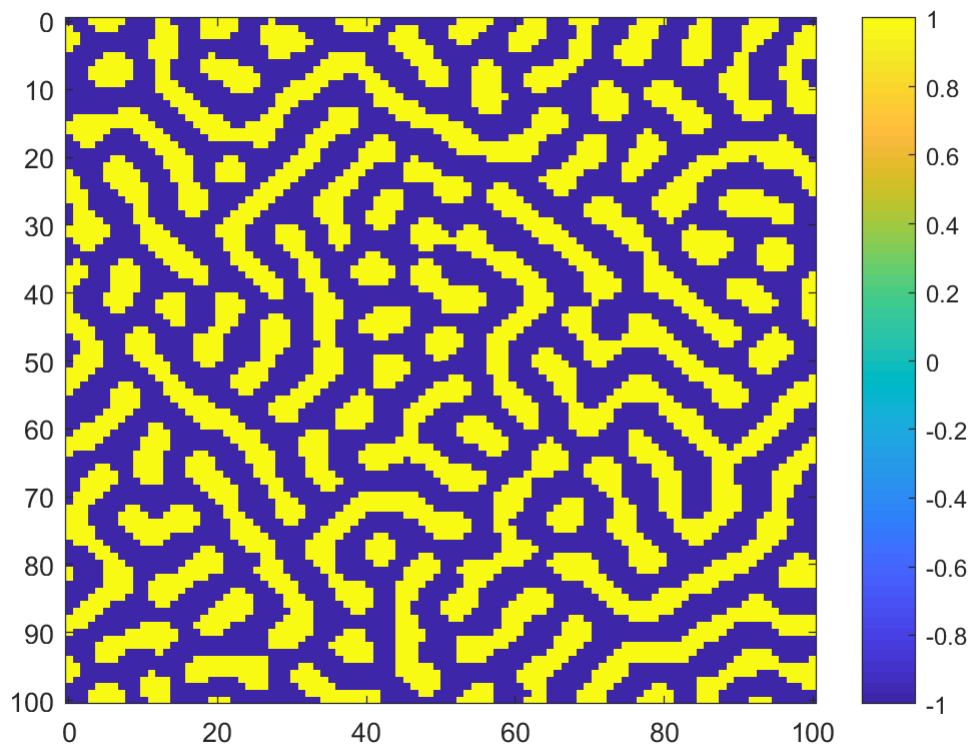


Figure 3:  $h = -2.73$ ,  $R_1 = 1.91$ ,  $J_1 = 1.73$ ,  $R_2 = 5.30$ ,  $J_2 = -0.40$

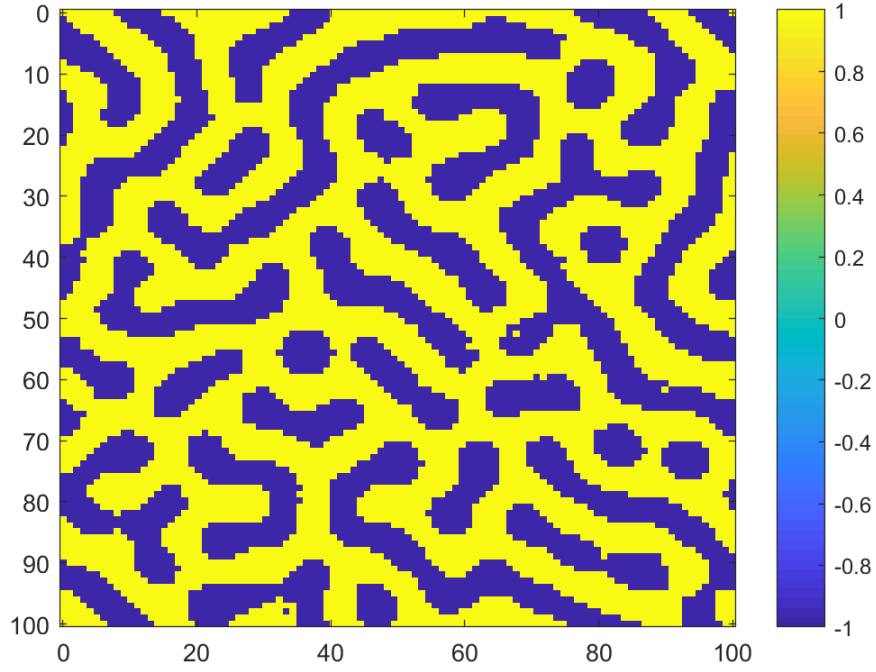


Figure 4:  $h = 4.93$ ,  $R_1 = 3.02$ ,  $J_1 = 2.62$ ,  $R_2 = 7.18$ ,  $J_2 = -0.82$

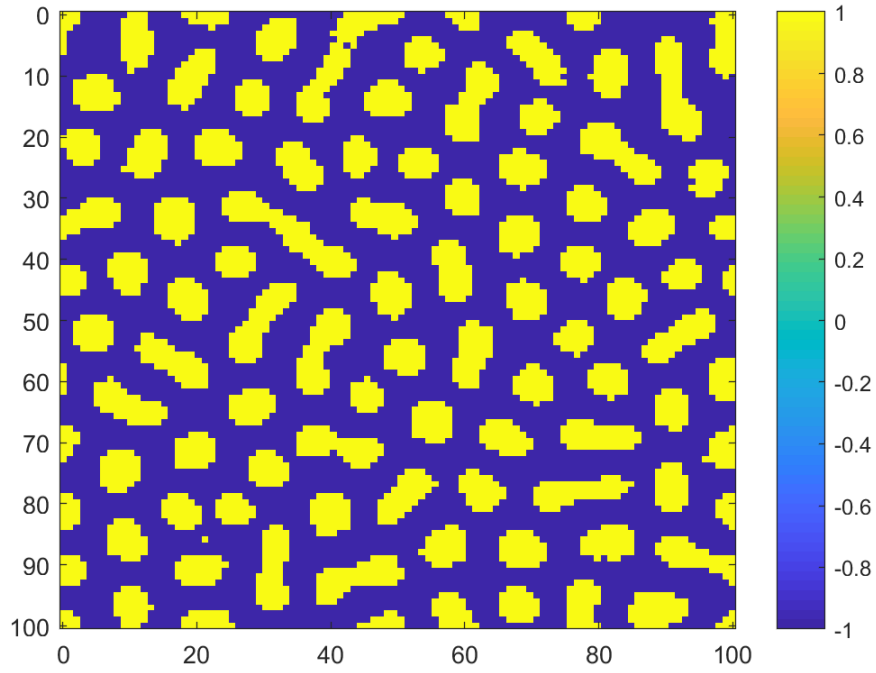


Figure 5:  $h = -13.47$ ,  $R_1 = 2.80$ ,  $J_1 = 2.43$ ,  $R_2 = 6.80$ ,  $J_2 = -0.73$

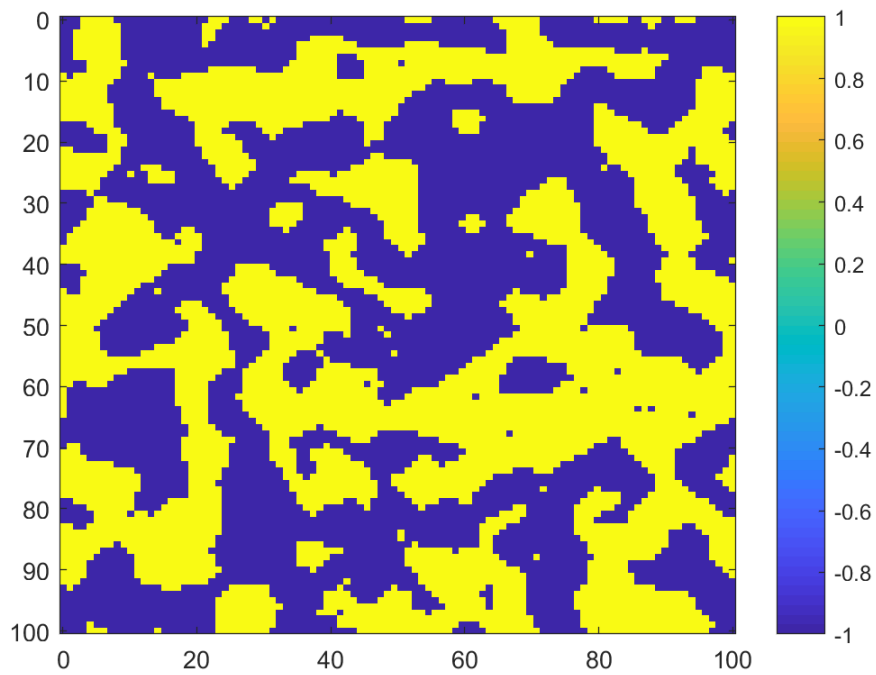


Figure 6: Original Ising Model with single radius:  $h = 0$ ,  $R_1 = 1.5$ ,  $J_1 = 10$

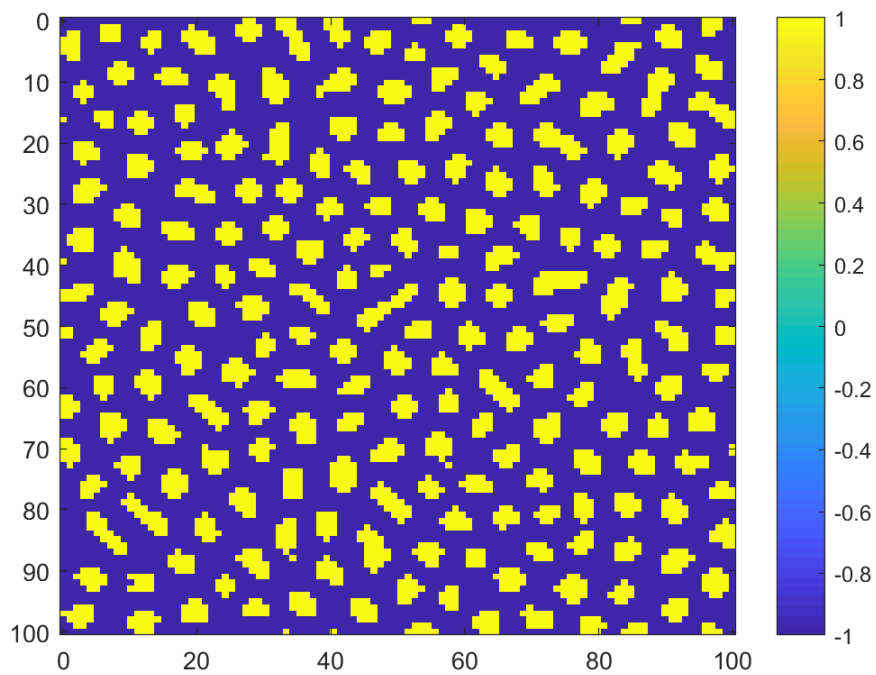


Figure 7:  $h = -3.79$ ,  $R_1 = 1.66$ ,  $J_1 = 1.53$ ,  $R_2 = 4.87$ ,  $J_2 = -0.30$

# Summary Paper

## A continuum approximation to an off-lattice individual-cell based model of cell migration and adhesion [1]

### 1 Introduction

Cellular migration is of great interest to modelers due to the large range of situations it is necessary to take into account. There have been a number of approaches to modeling this behavior, with the beginning models simplifying as much as possible. Later models increase in sophistication as the previous models are shown to be either insufficient in capturing the behavior desired or are inscrutable from the perspective of using the model to extract connections between the model and the microscopic cell-scale processes. The continuum approximation model proposed in this paper tries to address these two concerns first by identifying the deficiencies of previous pertinent models and then by extending those preceding models to compensate, resulting in a better continuum model.

### 2 Previous Models Discussion

The first model, a Mean Field Approximation, MFA, discussed revolves around a top-down approach, where a continuum-based model is derived from locally averaged properties such as spatial distributions of cell densities. This approach relies on either well-established physical laws or by intuition to describe the motion of the cells, but assumes that the statistical correlation of cell-cell interactions are negligible. Obviously this is problematic, as experimental data demonstrates clearly that there are strong interactions between cells and their nearby neighbors.

To try and capture this interaction between cells, IBM or "individual cell-based models," were created. This model treats each cell as a discrete entity, whereas the previous continuum model treated them as an averaged value. The first model using this discrete approach used a single node to describe the center of mass of each cell. These nodes are then placed on a lattice grid, and the motion is simulated using transition probabilities. These transition probabilities are where considerations such as volume exclusion and cell adhesion to either other cells or to the ECM are taken into account. The probability of vacating a node is lowered if there are neighboring cells (adhesion), but the issue with this approach is that the coarseness of lattice. The nodes representing the cells can be removed by more than the diameter of a cell, creating a non-lifelike condition that disallows the modelers from capturing correlations between the positions and velocities of the cells to each other. Attacking this issue of coarseness, the next logical step is to create numerous nodes to describe a single cell. A good example of this is the Cellular Potts Model. This approach allows the model to capture the previously unavailable correlation data, but does so at the cost of computational time, as the lattice can be thought of as being refined into a finer mesh. This computational cost induces investigation into off-lattice models. These models are comparable to the on-



lattice models in the sense that they too can either represent the cells as either centers of mass or a collection of nodes. The difference is that off-lattice better captures the behaviors of the cells even in the coarsest settings where there are single nodes assigned to each cell, with the positions of these cells or subcellular cell elements being governed by either stochastics or ordinary differential equations.

### 3 Setup

The discrete model is the more accurate representation of the behavior of the cells, but comes at high computational cost. Therefore, being able to connect the continuum models with the discrete model is ideal, as it would allow for the fast evaluation of variations of the parameter space. A discrete model is introduced using an off-lattice IBM based on Langevin equations to begin this process of connection. This model captures the adhesion and migration of the cells, and then is used to derive the governing equations for the associated probability distribution functions, which form a hierarchy of  $N$  partial integro-differential equations, where  $N$  denotes the number of cells in the system. The discrete model is run multiple times,  $10^5$  times to be exact, to create an "ensemble average" distribution, which creates numerical results by averaging the individual values across a large number of runs. This ensemble average is compared against the results generated from two different continuum approximations, the MFA and the Kirkwood Superposition approximation, KSA.

The continuum models are introduced as a strategy to try and truncate the hierarchy of integro-differential equations generated in the IBM model to a manageable amount. This is achieved by comparing the results generated from the IBM against those of the continuum models to see if there is sufficient connection to justify using the continuum models as a tool for rapid investigation of the parameter space.

### 4 IBM Construction and Discussion

The IBM used to capture cell movement and adhesion follows the model developed by Newman and Grima:

$$\dot{x}_i = \xi_i + \alpha \nabla_i \phi.$$

The  $\xi$  term encompasses noise,  $\alpha$  represents the chemotactic attitude, with positive indicating aggregation and negative, repulsion. The gradient of  $\phi$ , the chemical concentration, is evaluated at the cell's current position. The IBM model assumes that the cells are point masses, and is constructed using Newton's second law,

$$m_i \frac{d^2 x_i}{dt^2} = F_i^{visc} + \sum_{k, k \neq i} F_{i,k}^{int} + \xi_i.$$

Here,  $m_i$  represents the mass of cell  $i$ ,  $x_i$  the position,  $F_{i,k}^{int}$  the force generated by the interactions between cells  $i$  and  $k$ ,  $F_i^{visc}$  the viscous force acting on cell  $i$ , and  $\xi_i$  the stochastic force representing self-propulsion. The  $\xi$  values that are used in this model are sampled from a Gaussian distribution with zero mean and zero auto-correlation as follows:

$$\langle \xi_i(t), \xi_j(t') \rangle = 2D\delta_{i,j}\delta(t - t').$$

Note that the  $D$  in the Gaussian distribution represents the constant of proportionality, which corresponds to the cell diffusion coefficient. This allows for the assumption that cells will move randomly when isolated, but will experience mechanical interactions with their neighbors if they are proximally relevant. Additionally, the viscous substrate acting on cell  $i$  with a drag force can be assumed to be proportional to the cell's velocity, with value  $\mu$ , which can thus be ignored. This is acceptable because there is a low Reynolds number at the scale of the cellular length, which implies that the inertial forces  $F^{visc}$  can be neglected. There are further assumptions that the interaction forces can be written as

$$F_{i,k}^{int} = \hat{F}_0 F(\|x_i - x_j\|) \frac{x_i - x_j}{\|x_i - x_j\|}.$$

Here,  $F$  is a scalar function of the separation between pairs of cells, and  $\hat{F}_0$  contains the dimensional proportionality constant that captures the various mechanical forces. This term is particularly important as this term includes the cell-cell adhesion and cell resistance to deformation. This leads to the final equation

$$\frac{dx_i}{dt} = F_0 \sum_{k,k \neq i} F(\|x_i - x_j\|) \frac{x_i - x_j}{\|x_i - x_j\|} + \xi_i,$$

where  $F_0 = \hat{F}_0/\mu$ . For the sake of accuracy, non-linear force laws based on phenomenological principles is adopted, namely the Morse Potential, to describe  $F$ , as

$$F(r) = \begin{cases} 2(\exp(-2a(r - r_c)) - \exp(-a(r - r_c))), & r < \sigma r_c \\ 0, & r \geq \sigma r_c \end{cases}$$

$F(r)$  contains both attraction representing cell-cell adhesion, ( $r > r_c$ ), and repulsion representing resistance to cellular deformation, ( $r \geq r_c$ ). The mechanical interactions between the cell is restricted by  $\sigma r_c$ , where  $\sigma$  is the distance at which cells may interact, given in terms of multiples of cell-radii. The longer range interactions, biologically relevant due to motility protrusions such as filopodia, may be ignored as the protrusions are typically on the length scale of one cell radius, which falls below the  $\sigma$  threshold. The range that is considered for this model is  $1 < \sigma \leq 3$ , where  $\sigma \leq 1$  generates repulsive forces only.

Additional considerations include the nondimensionalization of length and time as follows:

$$x^* = \frac{x}{L}, t^* = \frac{t}{L^2/D}.$$

Therefore the final equations which govern motion for this model are

$$\frac{dx_i^*}{dt} = F_0^* \sum_{k,k \neq i} F(\|x_i^* - x_j^*\|) \frac{x_i^* - x_j^*}{\|x_i^* - x_j^*\|} + \xi_i^*$$

$$< \xi_i^*(t), \xi_j^*(t') > = 2D\delta_{i,j}\delta(t^* - t'^*)$$

and

$$F^*(r) = \begin{cases} 2(\exp(-2a^*(r^* - \epsilon)/\epsilon) - \exp(-a^*(r^* - \epsilon)/\epsilon))/\epsilon, & r < \sigma r_c \\ 0, & r \geq \sigma r_c, \end{cases}$$

with

$$F^*(r^*) = \frac{F_0 L}{D}, \epsilon = \frac{r_c}{L}, a^* = ar_c.$$

In actuality, the term  $\epsilon$  is considered to be extremely small, and thus the superscripts are dropped and work is done exclusively with the dimensionless model.

## 5 One and two-cell probability density functions

The dimensionless IBM can now be represented using probability density functions, PDF. There is an assumption that the boundary is periodic, The one-cell PDF is defined as

$$P_i(\mathbf{x}, t) = \langle \delta(x - x_i(t)) \rangle,$$

the two-cell PDF as

$$P_i(\mathbf{x}, \mathbf{x}', t) = \langle \delta(x - x_i(t)) \delta(x' - x_j(t)) \rangle,$$

and thus with manipulation, we obtain

$$\frac{\partial P_i(x, t)}{\partial t} = -\nabla^2 P_i - F_0 \nabla \int F(\|x - x'\|) \frac{x - x'}{\|x - x'\|} p_2(x, x', t) dx'$$

and the therefore the final governing equations

$$\begin{aligned} \frac{\partial p_1(x, t)}{\partial t} &= \nabla^2 p_1(x, t) - F_0(N-1) \nabla \int F(\|x_i - x_j\|) \frac{x - x'}{\|x - x'\|} p_2(x, x', t) dx' \\ \frac{\partial p_2(x, x', t)}{\partial t} &= \nabla^2 p_2(x, x', t) - F_0 \left( \frac{\partial}{\partial x} - \frac{\partial}{\partial x'} \right) \times \left( F(\|x - x'\|) \frac{x - x'}{\|x - x'\|} p_2(x, x', t) \right) \\ &\quad - F_0(N-2) \frac{\partial}{\partial x} \int F(\|x - x''\|) \frac{x - x''}{\|x - x''\|} p_3(x, x', x'', t) dx'' \\ &\quad - F_0(N-2) \frac{\partial}{\partial x'} \int F(\|x' - x''\|) \frac{x' - x''}{\|x' - x''\|} p_3(x, x', x'', t) dx''. \end{aligned}$$

## 6 The Mean Field Approximation, "MFA"

The number of equations necessary to properly model the behavior of N cells requires N integro-differential equations, which is computationally undesirable. In adapting a closure method, the number of equations can be truncated to a more tractable problem. Truncating at the first level corresponds to writing down the density-density correlation function in terms of the cell density distribution. Precisely this means that

$$p_2(x, x', t) = p_1(x, t) p_1(x', t).$$

Therefore

$$\frac{\partial p_1(x, t)}{\partial t} = -\nabla^2 p_1(x, t) - F_0(N-1) \nabla \cdot (p_1(x, t) \phi(x, t),$$

with

$$\phi(x, t) = \int F(\|x - x'\|) \frac{x - x'}{\|x - x'\|} p_1(x', t) dx'.$$

## 7 The Kirkwood superposition approximation, "KSA"

The MFA appears to not provide adequate approximations to the cell-density distribution generated through the ensemble approximation given by the IBM model. This induces the development of an alternative closure relation, the Kirkwood Superposition Approximation. This takes the form of

$$p_3(x, x', x'', t) = \frac{p_2(x, x', t)p_2(x, x'', t)p_2(x', x'', t)}{p_1(x, t)p_1(x', t)p_1(x'', t)},$$

which is typically a good approximation in the limits of low densities and is a valuable tool to approximate off-lattice models of cells. The explicit formula for the KSA, analogous to the first governing equation of our IBM model is

$$\begin{aligned} \frac{\partial p_2(x, x', t)}{\partial t} &= \nabla^2 p_2(x, x', t) - F_0 \left( \frac{\partial}{\partial x} - \frac{\partial}{\partial x'} \right) \\ &\times \left( F(\|x - x'\|) \frac{x - x'}{\|x - x'\|} p_2(x, x', t) \right) - F_0(N - 2) \frac{\partial}{\partial x} \\ &\times \int F(\|x - x''\|) \frac{p_2(x, x', t)p_2(x, x'', t)p_2(x', x'', t)}{p_1(x, t)p_1(x', t)p_1(x'', t)} dx'' \\ &- F_0(N - 2) \left( \frac{\partial}{\partial x'} \int F(\|x' - x''\|) \times \frac{x' - x''}{\|x' - x''\|} \frac{p_2(x, x', t)p_2(x, x'', t)p_2(x', x'', t)}{p_1(x, t)p_1(x', t)p_1(x'', t)} dx'' \right) \end{aligned}$$

## 8 Numerical Methods and Discussion

For the purposes of comparison, the models are all restricted to one dimension. The comparisons are made between the estimates of cell-density and density-density correlation functions that are obtained from the IBM and those obtained from the MFA and KSA. As both the MFA and KSA models take the form of advection-diffusion, particularly the KSA, there is a consideration taken for the peaks that can naturally form in the density-density correlation functions in the  $x - x'$ . Therefore it becomes expedient to rewrite the equations for the using the center of mass  $\eta = x + x'$  and  $\xi = x - x'$ , where

$$\eta = \begin{cases} \xi & \text{for } 0 < \eta \leq 1, \\ 2 - \xi & \text{for } 1 < \eta \leq 2. \end{cases}$$

The MFA and KSA models were both run by first discretising the space, and then integrating the ODEs. This was compared with the IBM, where the Euler-Maruyama method is employed, with timesteps of size  $\Delta t = 10^{-7}$  and  $t \in [0, 10^{-2}]$ . There were 51 points taken from this model, at regular intervals.

The initial conditions for the IBM model are as follows:

$$p_0(x) = \frac{f(x)}{\int_0^1 f(x) dx}$$

and

$$f(x) = \frac{1}{2}(\tanh(\alpha(x - \theta)) + \tanh(\alpha(1 - \theta - x))),$$

with  $\theta = 0.2$  and  $\alpha = 30$  for all simulations. These conditions correspond to cells being densely packed in a one-dimensional environment. These results are compared to the density-density correlation function of the MFA and KSA by computing

$$E = \frac{1}{MT} \sum_{i,j} (p_1(i,j) - p_1^{ens}(i,j))^2.$$

Setting  $1 \leq i \leq M$  and  $1 \leq j \leq T$ , where  $T=51$ , the results depend on the beginning parameters, and thus the paper provides examples of the numerical results with different beginning parameters in graphical form to help a reader visualize the difference in variability.

Comparisons between the results of the IBM run  $10^5$  times against either the MFA or the KSA results in the understanding that the KSA model is superior to the MFA as a continuum approximation. The primary driving force for this conclusion results from the fact that the KSA develops the same peaks in the density-density correlation that are induced in the IBM, while the MFA is constantly disagreement. The problems with the KSA arise when intercellular forces are weak, and so the  $F_0$  is increased to make each of the peaks sharp compared to the results generated by the IBM. The differences between the MFA and the KSA are most easily found demonstrated by varying the sensing ranges ( $\sigma$ ), the intercellular force strength ( $F_0$ ), and relative densities  $\rho$ . The KSA is going to be much more robust in terms of being able to properly describe cell-cell interactions when there are both weak interacting cells and when there are high densities to consider. This is contrasted by the MFA, which is most useful when  $F_0$  is low. This same idea holds true with other parameter variations as well, indicating that the KSA is a preferable method for continuum approximation.

## 9 Summary

The KSA model is identified as being a much more relevant continuum approximation to the IBM models. This model is able to capture behaviors simulated in the IBM under a variety of conditions, exceeding the ability of the MFA to handle both low and high densities. Additionally, the KSA is able to incorporate other cell-level phenomena such as heptotaxis. The issues of the model arise when there are few cells that are motile (cancerous metastasis for example). Additionally, the computational challenges of validating the KSA model against the IBM model in higher dimensions requires a more refined approach to the numerical approximations of the KSA model. The computational requirements are going to be prohibitive as there are formulas arising from each dimension. Thus, the next step of research along these particular lines involves looking into the Johnson-Kendall-Roberts model, that incorporates adhesive elastic spheres. The most obvious area for improvement stems from the fact that the models all are supposing that the cells are represented by center of mass nodes. The KSA model is good to approximate the results of the IBM for both small and large cell population sizes, cell propulsion and mechanical forces. This is wonderful, as it allows the modeler to evaluate parameter variation without endeavoring to run  $10^5$  simulations on the IBM!

## References

- [1] Alistair M Middleton, Christian Fleck, and Ramon Grima. “A continuum approximation to an off-lattice individual-cell based model of cell migration and adhesion”. In: *Journal of Theoretical Biology* 359 (2014), pp. 220–232. issn: 00225193. DOI: 10.1016/j.jtbi.2014.06.011. URL: <http://dx.doi.org/10.1016/j.jtbi.2014.06.011><https://linkinghub.elsevier.com/retrieve/pii/S0022519314003543>.