

# Random Walks for Modelling Cell Populations

## VRES 2013-14

Morgan Wall  
n7532962

31 January 2014

## 1 Introduction

This report documents the tasks completed by Morgan Wall as part of his 2013-14 VRES scholarship with Dr. Matthew Simpson. The project covered the use of random walks in simulating cell motility, proliferation, and interaction in a variety of biological contexts. Furthermore, the derivation and analysis of partial differential equations to model equivalent processes was considered.

## 2 Simulating Random Walks in 2D

Random walk models are widely applicable to a variety of engineering contexts. Several models, either analysed or proposed by Simpson, were researched [7, 8]. Each model accounted for cell motility and proliferation, which were essential to the various biological processes being considered. The models varied based on the consideration of cell interactions via volume exclusion or by the proliferation mechanism used. The models covered were reproduced using MATLAB and the C programming language in conjunction with Intel's Math Kernel Libraries (MKL).

### 2.1 Methodology

The methodology applied for simulating cell processes mimics that of Simpson's prior work [7, 8]. Simpson considered applying a rectangular lattice with nodes positioned equidistantly. As such, agents can assume positions at  $(i, j)$ , where  $i, j \in \mathbb{N}$ . Each non-boundary node  $(i, j)$  therefore has the neighbouring nodes:  $(i, j + 1)$ ,  $(i, j - 1)$ ,  $(i + 1, j)$ , and  $(i - 1, j)$ . Random walk simulations are conducted on this lattice by initially placing agents at one or more nodes. Typically, a combination of periodic [2] and reflective boundary conditions [5] are imposed on the lattice.

For discrete cell simulations, random walks involve spatial and temporal parameters. Consequently, time must be discretised into a series of time steps. At the start of a time step, consider the case where there are  $N$  agents present. As per Simpson's approach,  $N$  agents are selected at random and each selection results in an agent being "updated".

Depending on the biological process being modelled, an “update” can include both motility and proliferation events. A motility event models cell movement and involves generating a random number from a uniform distribution in determining whether an agent attempts to move to an adjacent node. An adjacent node is consequently selected by generating another uniformly distributed random number  $S$  and...

## 2.2 Implementation

## 2.3 Results

In modelling a population of agents, Simpson derived partial differential equations by formulating an agent conservation statement for the average occupancy of a lattice site [8]. To demonstrate the reproduced model, simulations were conducted and compared against the following initial value problem (IVP):

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} \quad -\infty < x < \infty \quad (1)$$

$$C(x, 0) = \begin{cases} 0 & x < -h \\ C_0 - h & -h < x < h \\ 0 & x > h \end{cases} \quad -\infty < x < \infty \quad (2)$$

which has the analytic solution

$$C(x, t) = \frac{C_0}{2} \left( \operatorname{erf} \left( \frac{h-x}{2\sqrt{Dt}} \right) + \operatorname{erf} \left( \frac{h+x}{2\sqrt{Dt}} \right) \right). \quad (3)$$

This IVP was derived in [8] and is used to describe agent densities for a one-dimensional problem without proliferation. Density data was obtained from the reproduced model by performing simulations on a two-dimensional lattice with periodic boundary conditions and averaging the occupancy of each column over multiple identically-prepared realisations. As seen in Figure 1, the reproduced model matches the analytic solution closely as time elapses.

### 2.3.1 Additional Analysis

In exploring the various properties of the reproduced model, additional simulations were conducted for differing initial conditions and total realisations performed. As expected from (3), modifying the initial cell density merely scales the agent density curves (see Figure 2). Furthermore, as expected by the Law of Large Numbers, the accuracy of the reproduced model was dependent on the number of realisations averaged over [6]. As seen between Figures 1 and 3, an increase in the total realisations performed results in higher accuracy in the final solution.

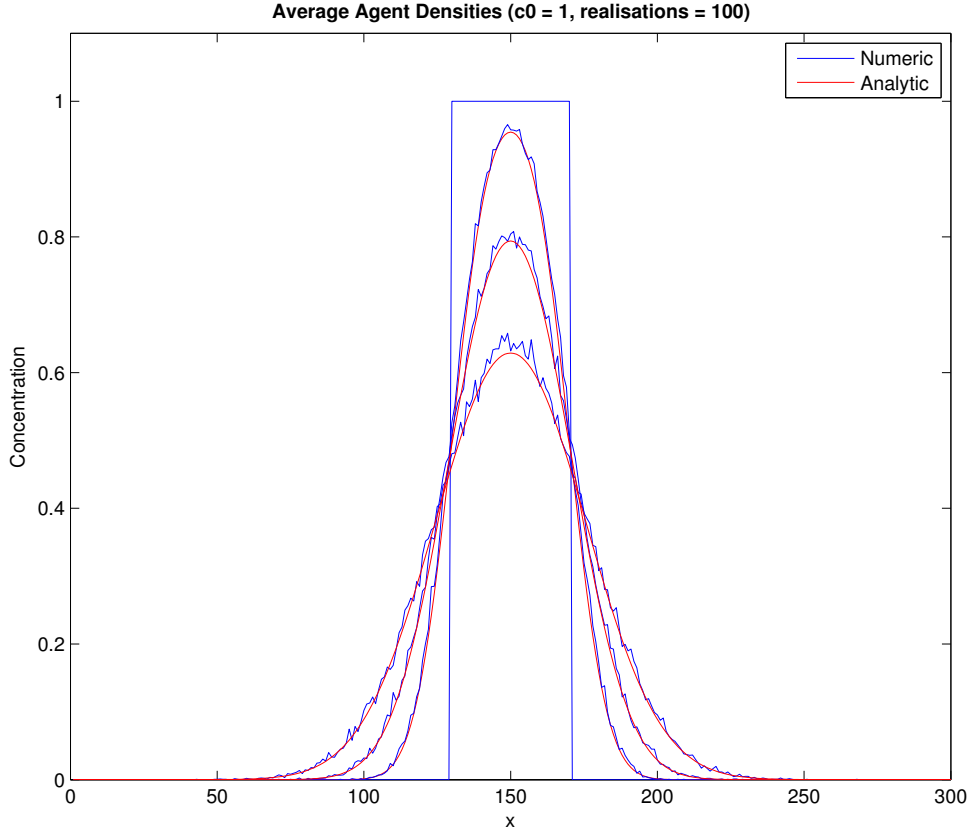


Figure 1: Comparison of agent density between the reproduced random walk model and the associated partial differential equation. Density data was obtained by averaging the occupancy of each column over 100 identically-prepared realisations satisfying (2) where  $C_0 = 1$ . Densities are given at  $t = 0, 200, 500$ , and  $1000$  and shows agent density diffusing across the lattice.

### 3 Random Lattices in 1D

In simulating cellular phenomena via random walk models, there exists three common methods for handling the spatial domain of interest: equally-spaced lattices, random lattices, and lattice free models. Lattices can be used under a variety of circumstances to discretise the spatial domain into a distinct set of nodes, which cells can consequently occupy. As discussed by Simpson [7], lattices can be constructed based off the expected size of a cell and the approximate distance they may move over a period of time. In constructing an equally-spaced lattice, nodes are spaced uniformly throughout a domain. Alternatively, nodes are positioned randomly or pseudorandomly within a domain in constructing a random lattice. Unfortunately, cells do not exhibit movement in a rigid, discrete manner, hence lattice-based models may not accurately simulate certain biological processes [3]. Lattice free models avoid this issue by not imposing a spatial discretisation on the simulations, which results in a major performance penalty.

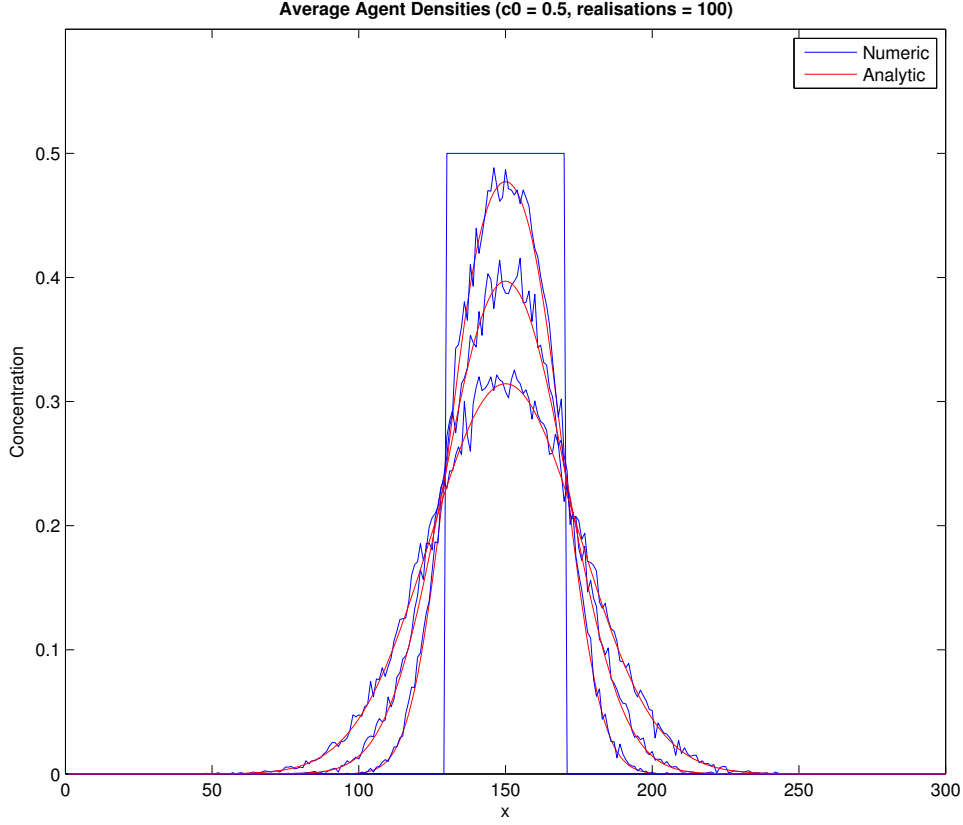


Figure 2: Comparison of agent density between the reproduced random walk model and the associated partial differential equation. Density data was obtained by averaging the occupancy of each column over 100 identically-prepared realisations satisfying (2) where  $C_0 = 0.5$ . Densities are given at  $t = 0, 200, 500$ , and 1000 and shows agent density diffusing across the lattice.

Research was conducted into the application of random lattices in simulating cell motility in one dimension. Two approaches were taken in constructing a random lattice: perturbing an equally-spaced lattice and generating node locations pseudorandomly. Under the former, each node position  $x_i$  was determined as follows:

$$x_i = y_i + X$$

where  $y_i$  is the location of the unperturbed node in the equally-spaced lattice and  $X$  is a variate sourced from a random distribution. Under the latter, each node position is given by:

$$x_i = Y.$$

where  $Y$  is a variate sourced from a random distribution. Despite both being random variates, the variance of the distributions from which  $\epsilon_i$  and  $\gamma_i$  were sourced from varied in comparative simulations. This distinction was established since it was quickly discovered that by imposing large perturbations on an existing lattice, node clustering and impractically large deltas between adjacent nodes can result. Furthermore, random variates were indexed

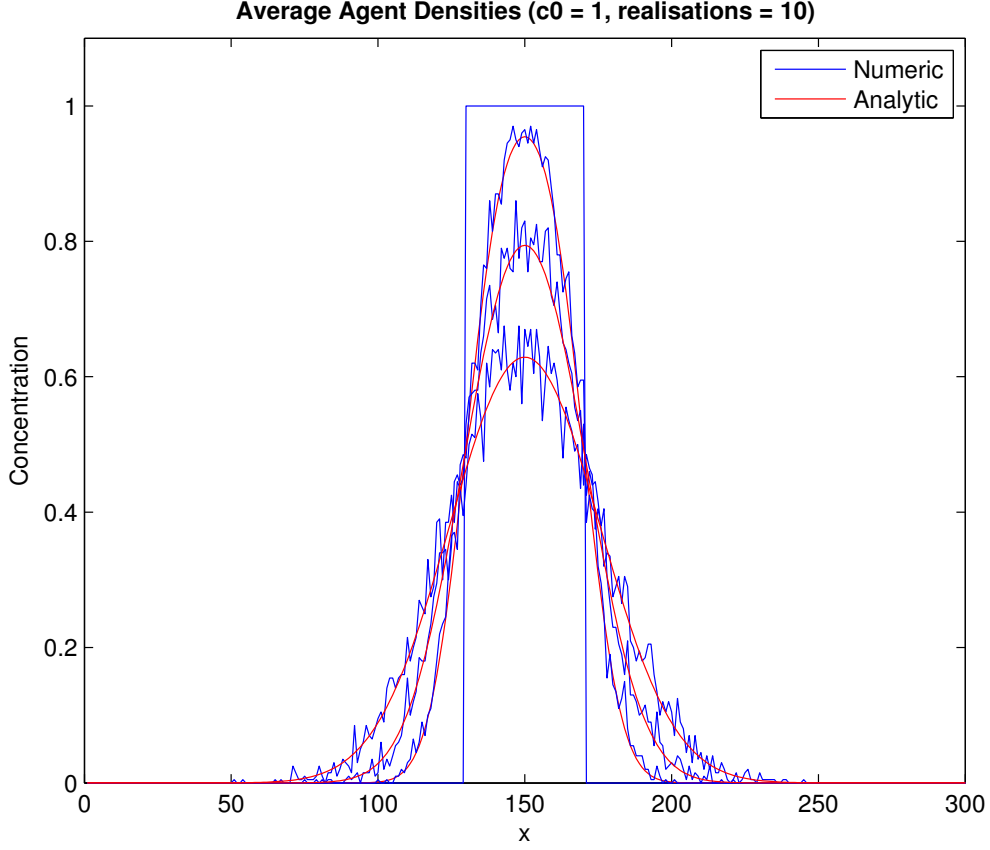


Figure 3: Comparison of agent density between the reproduced random walk model and the associated partial differential equation. Density data was obtained by averaging the occupancy of each column over 10 identically-prepared realisations satisfying (2) where  $C_0 = 1$ . Densities are given at  $t = 0, 200, 500$ , and  $1000$  and shows agent density diffusing across the lattice.

to reflect that pseudorandom numbers are retrieved from streams of values that, as a whole, exhibit properties of a specific statistical distribution. As such, in positioning the  $i^{\text{th}}$  node, the  $i^{\text{th}}$  value in the relevant stream was sourced. Unless otherwise specified, all random variates are sourced from a uniform distribution. In later simulations, pseudorandom numbers were sourced from a variety of distributions using Intel’s Vector Statistical Library (VSL).

### 3.1 Lattice Properties

Initially, several key characteristics of lattices were identified in determining the preferred method for generating random lattices. The deltas between adjacent nodes and the existence of node clusters were of particular interest in generating physically realistic lattices for simulating cellular phenomenon. A series of random lattices were constructed on a dimensionless 1D region of 20 length units in which 20 nodes would be positioned. For the perturbation method, the equally-spaced lattice consisted of 20 nodes spaced one

length unit apart with the first node residing at position zero.

Figure 4 depicts random lattices generated using the perturbation approach given

$$\epsilon_i \sim \text{Uniform}(-\Delta x, \Delta x) \quad (4)$$

where  $\Delta x$  is the maximum absolute perturbation allowed. These results quickly revealed that large perturbations are detrimental to the construction of random lattices since node clustering is evident. For example, consider the lattices where  $\Delta x = 0.1$  and  $\Delta x = 0.5$ . The former closely resembles an equally-spaced lattice whereas the latter exhibits node clustering around  $x = 6$  and  $x = 13$  and large unoccupied regions around  $x = 5$ . Such disparity in lattice spacing can result in erroneous simulations as cell motility can be subject to large variations in distance travelled per unit time. These findings are exacerbated as  $\Delta x$  is increased further and is can be shown mathematically by considering the expected value and variance of adjacent node deltas.

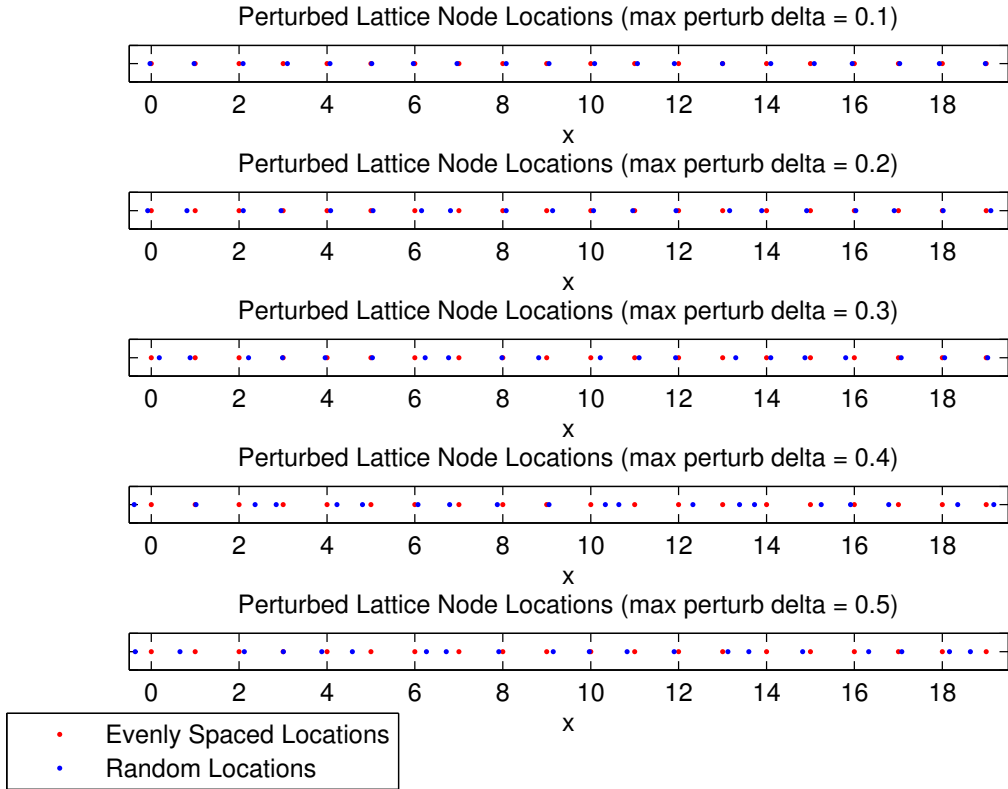


Figure 4

Consider a random lattice generated using the perturbing approach, where  $x_i$  is the position of the  $i^{\text{th}}$  node. Given the unperturbed lattice had a fixed node spacing  $\Delta x$  the expected node spacing is given by

$$\begin{aligned} E(x_i - x_{i-1}) &= E(y_i + \epsilon_i - y_{i-1} - \epsilon_{i-1}) \\ &= E(\Delta x + \epsilon_i - \epsilon_{i-1}). \end{aligned}$$

For generality, assume  $\epsilon_i$  and  $\epsilon_{i-1}$  are continuous random variables with probability density functions  $f(\alpha)$  and  $g(\beta)$ , respectively, then

$$\begin{aligned}
E(x_i - x_{i-1}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\Delta x + \alpha - \beta) f(\alpha) g(\beta) d\alpha d\beta \\
&= \Delta x \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\alpha) g(\beta) d\alpha d\beta \\
&\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \alpha f(\alpha) g(\beta) d\alpha d\beta \\
&\quad - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \beta f(\alpha) g(\beta) d\alpha d\beta \\
&= \Delta x + \int_{-\infty}^{\infty} g(\beta) d\beta \int_{-\infty}^{\infty} \alpha f(\alpha) d\alpha \\
&\quad - \int_{-\infty}^{\infty} f(\alpha) d\alpha \int_{-\infty}^{\infty} \beta g(\beta) d\beta \\
&= \Delta x + E(\epsilon_i) - E(\epsilon_{i-1}).
\end{aligned}$$

Hence, if  $\epsilon_i$  and  $\epsilon_{i-1}$  are independent and identically distributed continuous random variables, then

$$E(x_i - x_{i-1}) = \Delta x.$$

As such, if we were to investigate the average delta between adjacent nodes over many meshes, for example, then by the Law of Large Numbers, the expected delta is that of the equally-spaced lattice. This finding presupposes the expected value of  $\epsilon_i$  is zero. In this case, this finding does not describe the existence of node clustering and large deltas between nodes as seen previously. Instead, consider finding the variance of the same statistic as follows:

$$\begin{aligned}
\text{Var}(x_i - x_{i-1}) &= \text{Var}(y_i + \epsilon_i - y_{i-1} - \epsilon_{i-1}) \\
&= \text{Var}(\Delta x + \epsilon_i - \epsilon_{i-1}).
\end{aligned}$$

Making the same generality assumption for  $\epsilon_i$  and  $\epsilon_{i-1}$  as before and letting  $\Gamma = \Delta x + \epsilon_i - \epsilon_{i-1}$  be a continuous random variable gives

$$\text{Var}(x_i - x_{i-1}) = \text{Var}(\Gamma) = \int_{-\infty}^{\infty} (\gamma - E(\Gamma))^2 d\gamma$$

where

$$E(\Gamma) = \Delta x + E(\epsilon_i) - E(\epsilon_{i-1}).$$

Considering the case where  $\epsilon_i$  and  $\epsilon_{i-1}$  are independent and identically distributed continuous random variables, then

$$\begin{aligned}
\text{Var}(x_i - x_{i-1}) &= \int_{-\infty}^{\infty} (\gamma - \Delta x)^2 d\gamma \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\gamma - \Delta x)^2 d\gamma
\end{aligned}$$

## 3.2 Simulating Random Walks

### 3.2.1 Net Agent Displacement

## 4 Investigating Travelling Wave Solutions

### 4.1 Phase Plane Analysis of Fisher's Equation

As part of modelling cell invasion fronts, travelling wave solutions of partial differential equations was researched. The travelling wave solutions of Fisher's equation for population growth,

$$\frac{\partial U}{\partial t} = \nu \frac{\partial^2 U}{\partial y^2} + kU(1 - U), \quad (5)$$

was thoroughly analysed via [1]. To ease subsequent analysis, (5) can be nondimensionalised via the unit-reducing substitutions:

$$\tau = \alpha t \quad \text{and} \quad (6)$$

$$x = \beta y \quad (7)$$

where  $\alpha$  and  $\beta$  have units  $t^{-1}$  and  $m^{-1}$ , respectively. Applying the substitution in (6) and the chain rule, we obtain:

$$\begin{aligned} \frac{\partial U}{\partial \tau} \frac{d\tau}{dt} &= \nu \frac{\partial^2 U}{\partial y^2} + kU(1 - U) \\ \alpha \frac{\partial U}{\partial \tau} &= \nu \frac{\partial^2 U}{\partial y^2} + kU(1 - U) \end{aligned}$$

By inspection, it is clear that (5) can be nondimensionalised in time and simplified when  $\alpha = k$ . Hence,

$$\frac{\partial U}{\partial \tau} = \frac{\nu}{k} \frac{\partial^2 U}{\partial y^2} + U(1 - U) \quad (8)$$

Applying (7) in a similar manner gives

$$\frac{\partial U}{\partial \tau} = \frac{\beta^2 \nu}{k} \frac{\partial^2 U}{\partial x^2} + U(1 - U).$$

As such, (8) can be nondimensionalised in space and simplified when  $\beta = \sqrt{k/\nu}$ , which results in

$$\frac{\partial U}{\partial \tau} = \frac{\partial^2 U}{\partial x^2} + U(1 - U). \quad (9)$$

As discussed by Canosa [1], this nondimensionalised equation describes nonlinear population growth in one spatial dimension. Consequently, the modelled habitat (i.e. space) can only feasibly support a maximum population per unit length. Given the form of the logistic source term in (9), suppose that maximum population is unity. The following initial condition can therefore be applied:

$$0 \leq U(x, 0) \leq 1, \quad -\infty < x < \infty. \quad (10)$$



Canosa [1] was specifically interested solving (9) subject to (10) such that all the derivatives in  $x$  tended to zero as  $x \rightarrow \pm\infty$  coupled with the following far-field conditions:

$$\lim_{x \rightarrow -\infty} U(x, \tau) = 1 \quad \lim_{x \rightarrow \infty} U(x, \tau) = 0 \quad t \geq 0. \quad (11)$$

Any one-dimensional partial differential equations in space and time,  $U(x, \tau)$ , exhibiting travelling wave solutions supports the transformation

$$U(x, t) = U(x - c\tau) \equiv U(s) \quad (12)$$

which merges both time and space variables to provide a “standing wave” solution where the spatial origin is locked at the peak of the travelling wave moving at speed  $c$ . As such, the spatial coordinate system can be seen to translate in time in moving with the peak of the wave. Applying (12) with the chain rule gives

$$-c \frac{dU}{ds} = \frac{d^2U}{ds^2} + U(1 - U)$$

which has reduced Fisher’s equation into a second order non-linear partial differential equation. This equation can be further simplified into the following system of first order differential equations

$$\frac{dU}{ds} = V(s) \quad (13)$$

$$\frac{dV}{ds} = U^2 - U - cV. \quad (14)$$

Phase plane analysis was conducted on (13) and (14) using pplane8 for MATLAB [4].

Fisher originally discovered that (5) has travelling wave solutions where  $c \geq 2$ . This fact can be verified by solving the eigenvalue problem associated with (13) and (14) to determine the nature of any critical points in the phase plane. Given  $c \geq 2$  it can be shown that,  $(U, V) = (0, 0)$  is a stable node and  $(U, V) = (1, 0)$  is a saddle point. Conversely, given  $c < 2$ ,  $(U, V) = (0, 0)$  is a stable spiral. The phase planes seen in Figures 5 and 6 illustrate these results for specific values of  $c$ . Figure 5 exhibits the physically impossible solutions for  $c < 2$  as discussed by Canosa [1]. Note how certain phase trajectories can lead to negative values of  $U$ , which is physically impractical in modelling population growth.

## 4.2 Solving Fisher’s Equation

In order to better understand the travelling wave solutions to Fisher’s equation, the following IBVP was solved numerically using MATLAB’s pdepe function:

$$\begin{aligned} \frac{\partial U}{\partial \tau} &= \frac{\partial^2 U}{\partial x^2} + U(1 - U) & 0 \leq x \leq 100 & \quad t \geq 0 \\ U(0, \tau) &= 1 & \tau \geq 0 \\ U(100, \tau) &= 0 & \tau \geq 0 \\ U(x, 0) &= \begin{cases} 1 & x \leq 50 \\ 0 & \text{otherwise} \end{cases} & 0 \leq x \leq 100. \end{aligned}$$

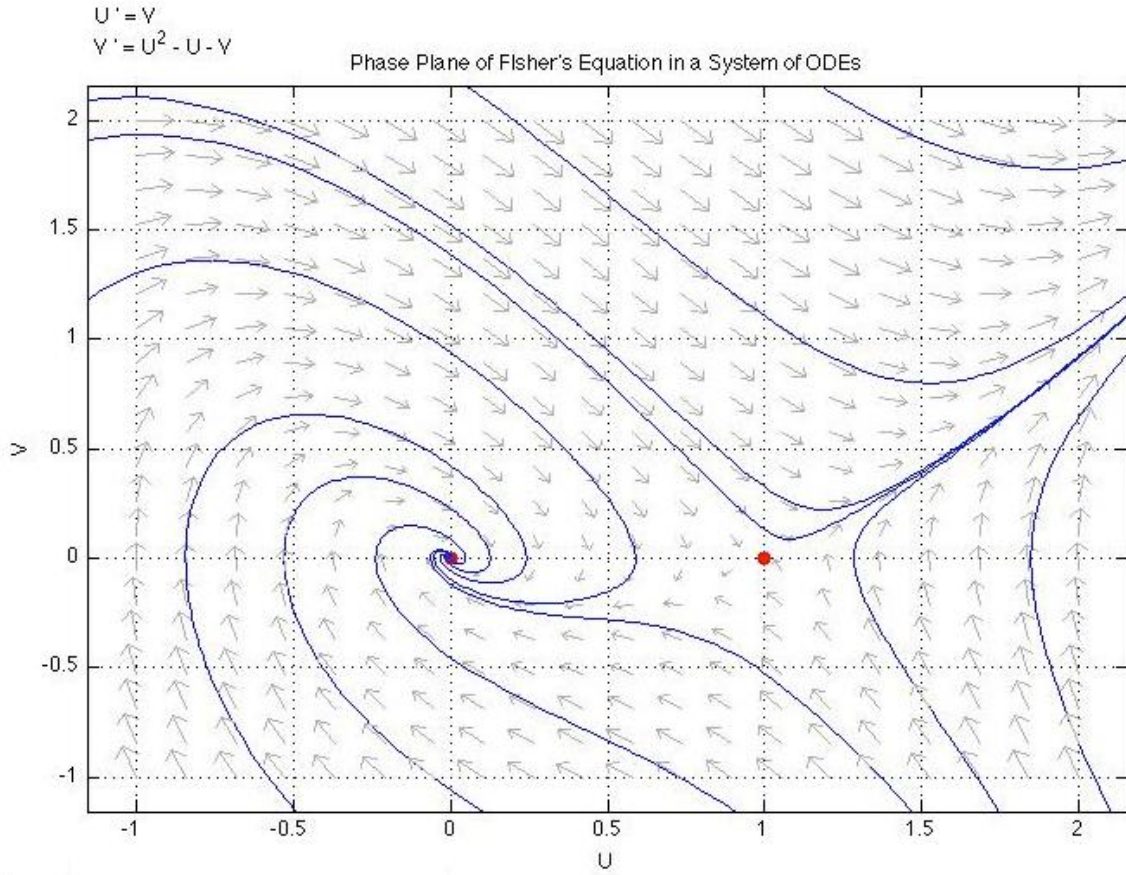


Figure 5:

The solutions, for varying timesteps, can be seen in Figure 7. From the figure it is clear that a static wavefront is formed which moves throughout space at a constant speed. This wavefront does not hold at the boundaries in order to satisfy the conditions imposed there. As a side note, the solutions did exhibit monotonicity issues. After discussion with Simpson, it was determined that this is due to MATLAB's `pdepe` function using the method of lines, which can result in physically unrealistic solutions. It was decided that alternative means for solving Fisher's equation would be pursued as a result.

Solutions to Fisher's equation were investigated for various initial conditions to observe how the logistic source term reacts to overpopulated and underpopulated regions. As expected, provided  $U \geq 0$ , the solutions continue to form a travelling wavefront. This is due to the logistic source term compensating for regions which do not satisfy the carrying capacity of unity. Likewise, in the event that  $U < 0$ , the solutions diverge to  $-\infty$ . It is clear that the logistic source term plays an integral role in forming travelling wave solutions of a specific amplitude.

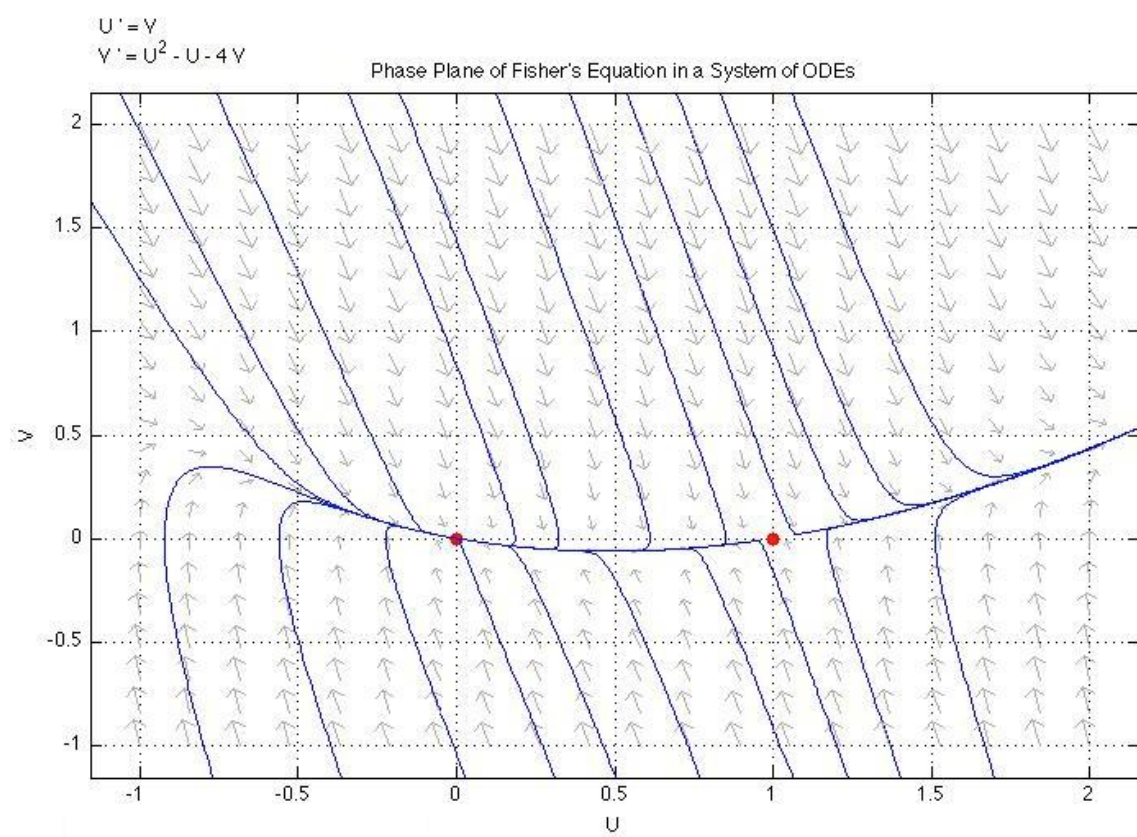


Figure 6:

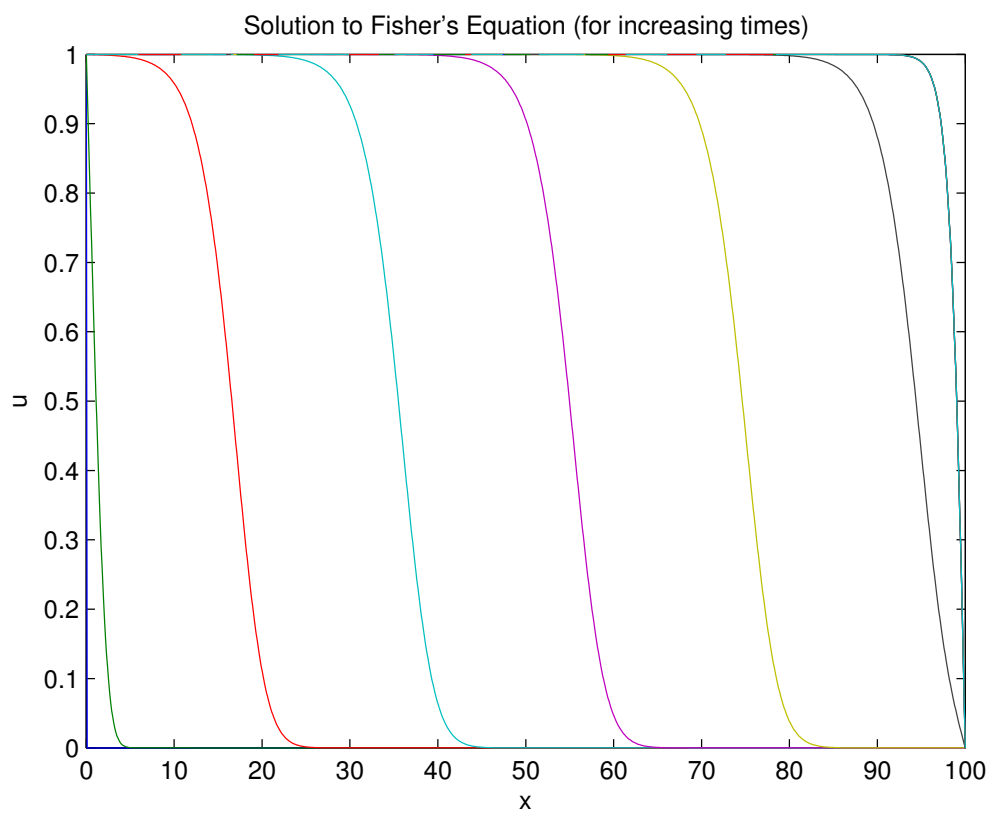


Figure 7: [insert figure description]

## References

- [1] José Canosa. “On a nonlinear diffusion equation describing population growth”. In: *IBM Journal of Research and Development* 17.4 (1973), pp. 307–313.
- [2] *Periodic Boundary Conditions*. Virtual Institute of Nano Films. 2012. URL: <http://www.mbnexplorer.com/users-guide/6-numerical-methods-interatomic-interactions/63-boundary-conditions/632-periodic>.
- [3] Michael J Plank and Matthew J Simpson. “Lattice-Free Models of Cell Invasion: Discrete Simulations and Travelling Waves”. In: *Bulletin of mathematical biology* 75.11 (2013), pp. 2150–2166.
- [4] John C Polking. *ODE Software for MATLAB*. Rice University. Apr. 2002. URL: <http://math.rice.edu/~dfield/>.
- [5] *Reflective Boundary Conditions*. Virtual Institute of Nano Films. 2012. URL: <http://www.mbnexplorer.com/users-guide/6-numerical-methods-interatomic-interactions/63-boundary-conditions/631-reflective>.
- [6] John Renze and Eric W Weisstein. *Law of Large Numbers*. MathWorld—A Wolfram Web Resource. URL: <http://mathworld.wolfram.com/LawofLargeNumbers.html>.
- [7] Matthew J Simpson, Kerry A Landman, and Barry D Hughes. “Cell invasion with proliferation mechanisms motivated by time-lapse data”. In: *Physica A: Statistical Mechanics and its Applications* 389.18 (2010), pp. 3779–3790.
- [8] Matthew Simpson, K Landman, and B Hughes. “Diffusing populations: ghosts or folks”. In: *Australasian Journal of Engineering Education* 15.2 (2009), pp. 59–68.