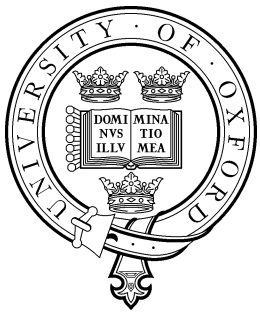
Spatial correlation models for cell populations



Deborah C. Markham

Keble College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2014

Abstract

Determining the emergent behaviour of a population from the interactions of its individuals is an ongoing challenge in the modelling of biological phenomena. Many classical models assume that the spatial location of each individual is independent of the locations of all other individuals. This mean-field assumption is not always realistic; in biological systems we frequently see clusters of individuals develop from uniform initial conditions. In this thesis, we explore situations in which the mean-field approximation is no longer valid for volume-excluding processes on a regular lattice. We provide methods which take into account the spatial correlations between lattice sites, thus more accurately reflecting the system's behaviour, and discuss methods which can provide information as to the validity of mean-field and other approximations.

Acknowledgements

I am grateful to my supervisors, Professors Ruth Baker, Matthew Simpson, Philip Maini and Eamonn Gaffney, for all their guidance throughout my D.Phil. In particular, thank you to Ruth for getting me into this field of research and being my primary supervisor. Thank you to Mat for hosting me in Brisbane in September 2013 for a very productive and enjoyable three weeks. In addition, thank you to Mat for the code provided for part of Chapter 3, and to him and Parvathi Haridas for providing the experimental images in this thesis.

Thank you also to my examiners, Professors Kieran Sharkey and Christian Yates, for their helpful comments and discussions.

This thesis would not have been possible were it not for the generous financial assistance provided over the past four years by the Oxford University Press (Clarendon Scholarship) and Keble College (Sloane-Robinson Award). My thanks go also to the Systems Biology Doctoral Training Centre, where I received my initial training.

I would like to give thanks for the funding I have received from various sources for my academic trips: the Mathematical Institute and Keble Association have both funded trips to Gothenburg and Knoxville for conferences, my trip to Australia was funded by the Royal Society International Exchange Scheme, and the National Cancer Institute funded a trip to the Moffitt Cancer Centre.

To David, Andrew, Kathy and Isla, thank you for everything. It's been great to have you all living in Oxford at some point in my D.Phil.

A huge thank you to my parents for having provided so many opportunities for me, and for their endless support over the years. I don't know how I would have managed without you; I am truly blessed to have you as my parents.

Finally, thank you to Stephen, my best friend and husband-to-be, for providing the daily D.Phil. essentials: tea, YouTube links, sympathy-not-advice and moral support!

Contents

1 Introduction **1**

1.1 Classical population models 2

1.1.1 Temporal models 2

1.1.2 Spatio-temporal models 5

1.2 Individual-based models 7

1.3 Deriving population models 8

1.3.1 The mean-field approximation 9

1.3.2 Comparing mean-field and averaged discrete results 12

1.3.3 Moment dynamics models 15

1.4 Conclusion 20

2 A single-species correlations model **24**

2.1 Introduction 25

2.2 The system 26

2.3 Discrete simulations 26

2.4 The mean-field approximation 28

2.4.1 Comparing the MFA and the IBM 31

2.5 Deriving the correlations ODEs 32

2.5.1 System closure 37

2.6 Solving the correlations ODE system 39

2.7 Results of the correlations ODE system 42
2.8 Conclusion 47

3 Comparing methods for modelling spreading cell fronts 48
3.1 Introduction 49
3.2 The methods 53
3.2.1 The discrete model 54
3.2.2 The mean-field approximation 56
3.2.3 Pair-wise approximation 58
3.2.4 One-hole approximation 62
3.3 Results 66
3.3.1 Transient behaviour 66
3.3.2 Asymptotic behaviour 70
3.4 Discussion 70
3.5 Conclusion 74

4 Simplifying the one-species correlations model 76
4.1 Introduction 77
4.2 Derivation in one dimension 79
4.3 Derivation in two dimensions 81
4.4 Derivation in three dimensions 84
4.5 Testing the PDE approximation 85
4.6 Insights from the PDE 86
4.7 Parameter investigations 91
4.7.1 Extinction region 92
4.8 Conclusion 94

5 A correlations model for the multi-species case **95**

5.1 Introduction 95

5.2 The general system 97

5.2.1 Equations for the N-species case 98

5.2.2 Deriving the MFA 100

5.2.3 The correlations approach 101

5.2.3.1 Deriving the correlations ODEs 101

5.2.3.2 Deriving the correlations PDEs 108

5.3 A specific case: two species 110

5.3.1 The discrete system 110

5.3.2 The MFA 112

5.3.3 The correlations PDEs 113

5.3.4 Systems undergoing proliferation and movement 115

5.3.5 Including death 120

5.3.6 Trading movement for proliferation 122

5.3.7 Break down of the correlations method 122

5.4 Conclusions and future work 126

6 Clustering measures **128**

6.1 Introduction 129

6.2 Agent coordination number method 130

6.2.1 One-species system 131

6.2.2 Two-species system 136

6.3 Spatial index method 140

6.3.1 One-species system 142

6.3.2 Two-species system 144

6.4 Spatial correlations method 147

6.4.1 One-species system 149

6.4.2 Two-species system 149
6.5 Comparing the clustering measures 150
6.6 Conclusions 155

7 Conclusion **157**
7.1 Future directions 161

Bibliography **165**

Chapter 1

Introduction

Mathematics is playing an increasingly crucial role in the biological sciences. In conjunction with experimental investigations, mathematical models have helped us gain insight into biological phenomena such as cancer [Byrne et al., 2006; Gatenby et al., 2006], wound healing [Dale et al., 1994; Maini et al., 2004a,b] and embryonic development [McLennan et al., 2012; Tomlin and Axelrod, 2007]. However, the accuracy of any theoretical predictions is highly dependent on the models chosen to represent a given system. If the underlying model assumptions are flawed, then the results produced by the model could lead to inaccurate predictions. This could have costly consequences, and may also result in experimentalists being hesitant about incorporating mathematics into further investigations. Moreover, if models are excessively complex, or computationally expensive to implement, their applications in biological scenarios will be greatly restricted. Thus we must understand the limitations of the models we are using, and be informed about their relative complexity, ensuring that we opt for models that are most suited to a given application. In this introductory chapter we provide an overview of some mathematical modelling techniques commonly used for biological applications, and examine their strengths and weaknesses. We begin by discussing two broad classes of population models: those in which the

population numbers are assumed to vary only in time and those in which they are assumed to vary in both space and time. We then turn to individual-based models, before finally considering methods that bridge the gap between descriptions at the individual and population levels.

1.1 Classical population models

Population models describe the behaviour of an entire group without explicity accounting for every individual. As biological populations often comprise many agents, being able to model their behaviour as a collective entity greatly reduces the complexity of the system, generally leading to simple and tractable models. Classically, we assume that biological populations can be modelled using standard physical principles such as the law of mass action and Fickian diffusion. Many population level models are deterministic: randomness plays no part at the scale of modelling. Thus, beginning from the same initial conditions, a deterministic model will give the same results upon repetition. The standard justification for using deterministic population models is that we expect random effects to have negligible consequences for large populations. Deterministic population models usually take the form of differential equations which can often be explored using standard analytical or numerical techniques. In this section, we consider some of the most commonly used deterministic population models, both with and without spatial dependence.

1.1.1 Temporal models

In certain systems, the behaviour of the population is assumed to be independent of its spatial distribution. Therefore, the equations describing population growth will only depend on time, leading to temporal ordinary differential equations (ODEs). The well-known logistic equation, first devised by Verhulst in the 19th century to

describe population growth [Verhulst, 1838], has the following form:

$$\frac{\mathrm{d}N}{\mathrm{d}t} = rN\left(1 - \frac{N}{K}\right), \tag{1.1.1}$$

where N is the number of individuals, r is the intrinsic growth rate, and K is the carrying capacity of the environment. This equation can be solved exactly to give:

$$N(t) = \frac{N_0 K e^{rt}}{K + N_0(e^{rt} - 1)}, \tag{1.1.2}$$

where N_0 is the value of N at $t = 0$. This solution is known as a logistic function, and it has a characteristic sigmoidal shape (see Figure 1.1). The number of individuals is often normalised by the carrying capacity, which allows us to examine the population density, giving the following form:

$$\frac{\mathrm{d}c}{\mathrm{d}t} = rc(1 - c), \tag{1.1.3}$$

where c is the population density, $c \in [0, 1]$. The logistic equation has been widely used in the mathematical modelling of population growth [Murray, 2002; Pielou, 1969]. For example, it is often used to describe bacterial population growth in microbiology experiments [Bailey and Ollis, 1986; Cabrero et al., 1998; Kacena et al., 1999], or to describe the behaviour of cancerous cell populations and tumour progression [Spratt et al., 1993; Vaidya and Alexandro, 1982]. This basic model has seen many extensions, including the incorporation of time delays [May, 1975; Murray, 2002] or modification of the growth functions [Tsoularis and Wallace, 2002].

Populations do not generally evolve in isolation, thus it is often necessary to incorporate interactions between different populations. If species do not directly interact, but compete for the same resources, we can couple logistic equations via the carrying capacity terms. Populations may also be directly interacting, and these interactions

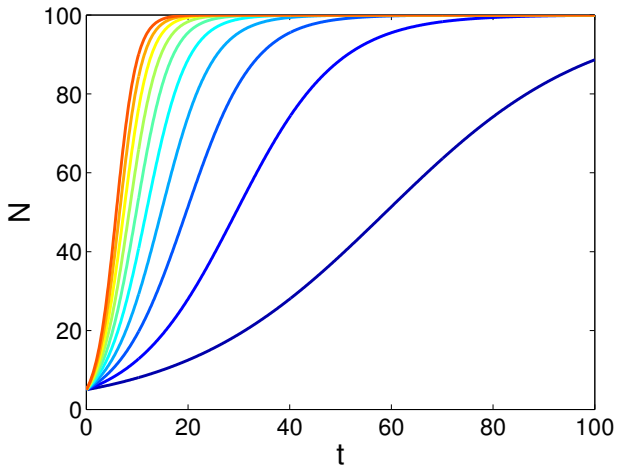


Figure 1.1: The solution, equation (1.1.2), to the logistic equation, equation (1.1.1), for a range of r values. In all cases, $K = 100$ and $N_0 = 5$. The growth rate, r, increases from 0.05 (blue) to 0.5 (red) in steps of 0.05. As r increases, the population evolves to its final state more rapidly. This steady-state is given by the carrying capacity, K, for all positive values of r.

can be explicitly included. In the early 20th century, Lotka [Lotka, 1925] and Volterra [Volterra, 1926] independently developed a pair of first-order non-linear ODEs for describing predator-prey interactions. Models of this form are straightforward to investigate using linear stability analysis and phase-plane techniques [Murray, 2002]. Another class of interacting models were developed to describe epidemics [Anderson and May, 1979; Kermack and McKendrick, 1991a,b,c; May and Anderson, 1979]. Known as SIR models, these have three compartments: susceptible (S), infected (I) and recovered (R). Three coupled non-linear ODEs are used to describe the evolution of the respective compartments. Models based on the SIR formulation can be used to examine epidemics, as well as strategies, such as vaccination programs, for reducing the impact of an infectious disease [Murray, 2002].

All the models we have discussed thus far are only suitable for the spatially invariant, or spatially averaged, case. However, it is often necessary to account for situations where the spatial distribution of agents plays a vital role. For example, wound healing relies on moving fronts of cells [Maini et al., 2004a,b], as does inva-

sion of healthy tissue by cancerous cells [Allred, 2010; Gatenby and Gawlinski, 1996; Swanson et al., 2003]. Phenomena in which the density dynamics depend on space can be modelled by using partial differential equations (PDEs) instead of ODEs, thus allowing the population to vary in space as well as time.

1.1.2 Spatio-temporal models

In cases where the population varies in both space and time, PDEs are typically used to describe the evolution of populations with respect to both variables. One of the most widely used spatio-temporal equations for describing biological populations is the Fisher-Kolmogorov (FK) equation, which combines logistic growth with linear diffusion. In one spatial dimension, this generally takes the following form:

$$\frac{\partial c}{\partial t} = rc(1-c) + D\frac{\partial^2 c}{\partial x^2}, \tag{1.1.4}$$

where $c \equiv c(x,t)$ is the population density, D is the diffusion coefficient and r is again the intrinsic growth rate. This equation can be shown to evolve to a constant speed travelling wave solution with a speed of $v = 2\sqrt{rD}$ [Murray, 2002], given initial conditions of the form

$$c(x,0) = c_0(x) \geq 0, \qquad c_0(x) = \left\{ \begin{array}{ll} 1, & x \leq x_1\,, \\ 0, & x \geq x_2\,, \end{array} \right\} \tag{1.1.5}$$

where $x_1 < x_2$ and $c_0(x)$ is continuous in $x_1 < x < x_2$. In Figure 1.2, we see an example of solutions to the FK equation at various time points for the case where $D = r = 1$, with step initial conditions and zero-flux boundary conditions. As expected, we see a travelling wave solution with $v = 2$.

The FK equation has many applications in biological settings, such as wound-healing [Maini et al., 2004a,b], tumour growth [Allred, 2010; Swanson et al., 2003]

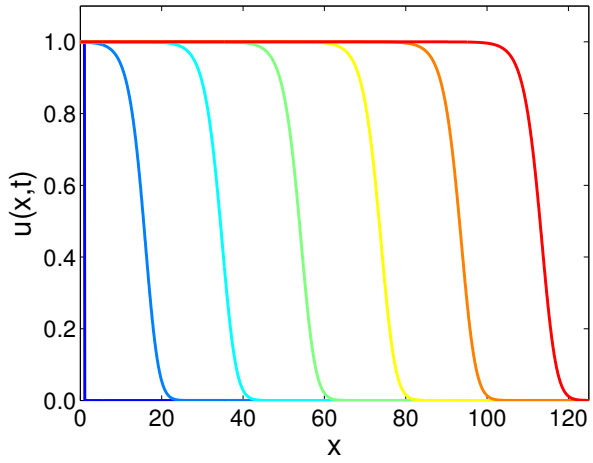


Figure 1.2: The solution of the FK equation, equation (1.1.4), is a travelling wave. In this case, $D = r = 1$ and we have shown the density at times from $t = 0$ (blue) to $t = 60$ (red) in steps of 10. Equation (1.1.4) is solved numerically using the built-in `pdepe` function in MATLAB with zero-flux boundary conditions.

and ecology [Shigesada and Kawasaki, 1997]. FK equations can be coupled to each other, and we can also allow reactions between the species. Reaction-diffusion models are frequently used to describe pattern formation [Liu et al., 2006; Turing, 1952] as well as the invasion of healthy tissue by cancerous cells [Gatenby and Gawlinski, 1996; Gatenby et al., 2006].

Whilst continuum descriptions of populations can be analytically and numerically tractable, there is the concern that behaviour at the individual level is being neglected. The behaviour of a population is a direct result of the interactions of its individual entities, thus phenomenological population models may not always be an accurate representation of the collective behaviour. For example, our phenomenological models ignore the volume-excluding effects of the individual agents. It is essential that, in addition to our population models, we also have models that explicitly account for the behaviour of individuals. From these, it can be determined if our phenomenological models are providing suitable approximations that capture the influence of the behaviour at the individual level.

1.2 Individual-based models

With the advances in computational power over the past few decades, we have become able to perform increasingly intensive computational simulations. This allows us to explictly account for every individual in a population, and thus examine the emergent population behaviour. In an individual-based model (IBM), each individual agent has a set of rules which govern its behaviour, and an algorithm is used to update the behaviour of the system [Grimm, 1999]. For example, the Gillespie algorithm [Gillespie, 1977], a type of dynamic Monte Carlo method, is frequently used. Volume-exclusion can be explicitly encoded into the system by making it impossible for agents to occupy the same physical space [Simpson et al., 2009b, 2010b]. IBMs can be on-lattice, in which case agents occupy positions on a regular lattice (usually comprised of squares [Smallbone et al., 2007] or hexagons [Alarcón et al., 2003]), and move between these specific sites [Ermentrout and Edelstein-Keshet, 1993]. Off-lattice IBMs no longer constrain the agents to locations on the lattice but allow them to occupy any location in space, which is more realistic but more challenging to implement [Galle et al., 2006; Johnston et al., 2013; Plank and Simpson, 2012, 2013]. Generally, IBMs are stochastic: multiple simulations will not produce identical results. Therefore, multiple simulations must be performed if we would like to obtain the average behaviour or range of behaviours [Ermentrout and Edelstein-Keshet, 1993].

IBMs have been applied to many biological phenomena [Codling et al., 2008], including predator-prey models [Cuddington and Yodzis, 2000; De Roos et al., 1991], angiogenesis [Owen et al., 2009; Shirinifard et al., 2009] and wound healing [Dallon et al., 2000; Mi et al., 2007]. We can extract population traits directly from the individual interactions, thus allowing us to compare the predictions from IBMs to those obtained from the models we would use to phenomenologically describe the same system. IBMs have been found to support the results of phenomenological models in some cases, but to contradict them in other cases [Durrett and Levin,

1994]. This disagreement between phenomenological descriptions and IBMs casts doubt on the accuracy of the phenomenological models when it comes to capturing behaviours that result from the individual agents.

Whilst IBMs allow investigation of the emergent population behaviour from its individual constituents, these models are computationally expensive as the behaviour of every agent in the system needs to be explicitly accounted for. Moreover, we need to average over many simulations to understand the average population trends, which can be time-consuming. In contrast to phenomenological models, it is also not usually possible to perform any mathematical analysis on IBMs, which can help gain valuable insight into the system. These issues can potentially be overcome if we are able to derive tractable population models directly from the individual behaviour. Linking IBMs and continuum models can help us understand how different models may be related, and confirm the underlying assumptions that are present in phenomenological models. If we derive models at the population level directly from the individual behaviour, we can perhaps identify simple models which still capture the underlying individual behaviour, without incurring the computational cost associated with many IBMs.

1.3 Deriving population models

Phenomenological modelling of populations is generally tractable numerically, and sometimes analytically, however these models may not capture the influence of behaviour at the individual level. By using IBMs, we can retain descriptions of the individuals. However, we must generally sacrifice the possibility of any analytical insight and incur a large computational cost. Ideally, we would like a modelling approach that lies somewhere between these two extremes, whereby we retain a link to the individual behaviour but without such a cost. Recently, a number of methods have

been developed aimed at deriving continuum models directly from IBMs [Baker and Simpson, 2010; Dodd and Ferguson, 2009; Dyson et al., 2012; Pennington et al., 2011; Simpson et al., 2009b; Yates, 2014; Yates et al., 2012]. Naturally, various assumptions need to be made along the way. We begin with the simplest of these methods, the mean-field approximation (MFA), applying it to a volume-exluding system.

1.3.1 The mean-field approximation

The MFA can be used to obtain a deterministic description of an IBM using master equations and assuming that the spatial locations of agents are independent. For example, let us begin with agents initially distributed uniformly at random on a one-dimensional lattice with lattice spacing Δ. Let us assume that these agents proliferate and migrate at rates of P_p and P_m per unit time, respectively. Further, we assume that proliferation events place a daughter cell in a randomly chosen neighbouring site. Similarly, migration events allow agents to move to a randomly chosen neighbouring site. The occupancy of site l is given by σ_l where $\sigma_l = A$ if site l is occupied by species A, otherwise it is 0. If the chosen neighbour is occupied, then the event does not occur. From these rules we can write a master equation for the evolution of the probability of occupancy, $\mathbb{P}(A_l)$, of a lattice site, l

$$\begin{aligned}\mathbb{P}(A_l(t+\delta t)) - \mathbb{P}(A_l(t)) = & \underbrace{\frac{P_m}{2}\mathbb{P}(A_{l-1}, 0_l)\delta t}_{\text{movement into site } l \text{ from the left}} + \underbrace{\frac{P_m}{2}\mathbb{P}(0_l, A_{l+1})\delta t}_{\text{movement into site } l \text{ from the right}} \\ & - \underbrace{\frac{P_m}{2}\mathbb{P}(0_{l-1}, A_l)\delta t}_{\text{movement out of site } l \text{ to the left}} - \underbrace{\frac{P_m}{2}\mathbb{P}(A_l, 0_{l+1})\delta t}_{\text{movement out of site } l \text{ to the right}} \\ & + \underbrace{\frac{P_p}{2}\mathbb{P}(A_{l-1}, 0_l)\delta t}_{\text{proliferation into } l \text{ from the left}} + \underbrace{\frac{P_p}{2}\mathbb{P}(0_l, A_{l+1})\delta t}_{\text{proliferation into } l \text{ from the right}},\end{aligned} \tag{1.3.1}$$

where $\mathbb{P}(\sigma_{l-1}, \sigma_l)$ describes the joint probability that sites $l-1$ and l have occupancies

σ_{l-1} and σ_l, respectively. In the MFA, we assume that the occupancy probabilities of sites are independent, thus

$$\mathbb{P}(\sigma_{l-1}, \sigma_l) = \mathbb{P}(\sigma_{l-1})\mathbb{P}(\sigma_l). \tag{1.3.2}$$

Setting $\mathbb{P}(A_l(t)) \equiv c_l(t)$, and assuming lattice site independence, equation (1.3.1) becomes

$$\begin{aligned} c_l(t+\delta t)) - c_l(t) = & \underbrace{\frac{P_m}{2} c_{l-1}(t)[1-c_l(t)]\delta t}_{\text{movement into site } l \text{ from the left}} + \underbrace{\frac{P_m}{2} c_{l+1}(t)[1-c_l(t)]\delta t}_{\text{movement into site } l \text{ from the right}} \\ & - \underbrace{\frac{P_m}{2} c_l(t)[1-c_{l-1}(t)]\delta t}_{\text{movement out of site } l \text{ to the left}} - \underbrace{\frac{P_m}{2} c_l(t)[1-c_{l+1}(t)]\delta t}_{\text{movement out of site } l \text{ to the right}} \\ & + \underbrace{\frac{P_p}{2} c_{l-1}(t)[1-c_l(t)]\delta t}_{\text{proliferation into } l \text{ from the left}} + \underbrace{\frac{P_p}{2} c_{l+1}(t)[1-c_l(t)]\delta t}_{\text{proliferation into } l \text{ from the right}} . \end{aligned} \tag{1.3.3}$$

As the agents are uniformly distributed initially, we expect the occupancy probabilities of each lattice site to be identical on average, thus for all values of i, $c_i = c$. We divide through by δt, then take the limit $\delta t \to 0$, to obtain

$$\frac{\mathrm{d}c}{\mathrm{d}t} = P_p c(1-c). \tag{1.3.4}$$

Comparing equation (1.3.4) with equation (1.1.3), we note that the MFA in this case gives us the logistic population density equation where $r = P_p$. Thus, in this case, the phenomenological logistic description is implicitly based on the underlying assumptions of the MFA applied to this discrete system.

We now consider non-uniform initial conditions, with agents undergoing the same events again on a 1D lattice. We once more obtain equation (1.3.3), but now we cannot assume that all $c_i = c$. Instead we Taylor expand in our spatial index, keeping

terms up to second order, around c_l. Simplifying we obtain

$$\frac{\mathrm{d}c_l}{\mathrm{d}t} = \frac{P_m\Delta^2}{2}\frac{\partial^2 c_l}{\partial x^2} + \frac{P_p}{2}(1-c_l)\left(2c_l + \Delta^2\frac{\partial^2 c_l}{\partial x^2}\right), \tag{1.3.5}$$

where x is our spatial variable. Taking the limit as $\Delta \to 0$, allowing P_m to scale with $1/\Delta^2$ [Codling et al., 2008; Othmer et al., 1988; Simpson et al., 2009a, 2010a; Stevens and Othmer, 1997], we obtain

$$\frac{\partial c}{\partial t} = \frac{\hat{P}_m}{2}\frac{\partial^2 c}{\partial x^2} + P_p c(1-c), \tag{1.3.6}$$

where $c \equiv c(x,t)$ is the population density function, and $\hat{P}_m = P_m\Delta^2$. We notice that this equation has the same form as the Fisher equation, seen in equation (1.1.4), with $r = P_p$ and $D = \hat{P}_m/2$.

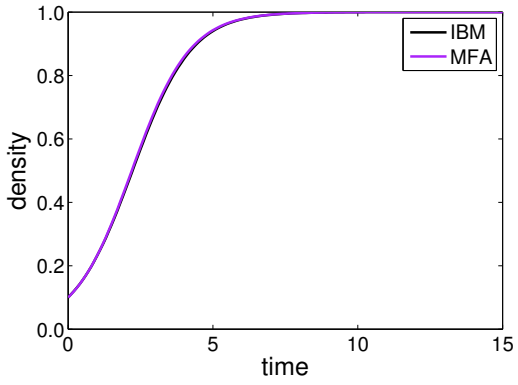
Using the MFA on these specific IBMs, we have recovered the phenomenological models we discussed in Section 1.1, both with and without spatial dependence. These models are frequently used in biological applications without rigorous derivation; the recovery of these models by this derivation supports their use for these specific systems when the underlying assumptions are satisfied. Different rules for the IBM will naturally lead to different equations, and the MFA is often used to scale up specific individual systems, generating results which in some circumstances agree with the individual behaviour [Codling et al., 2008; Deroulers et al., 2009; Sumpter and Broomhead, 2001]. However, the MFA is not necessarily a suitable approximation under all circumstances, which implies that our standard phenomenological models, and other derived mean-field models, may not always be a valid choice. To look at this in more detail, we examine the two mean-field models derived in this section in two broad areas of parameter space: a large P_p/P_m ratio and a small P_p/P_m ratio.

1.3.2 Comparing mean-field and averaged discrete results

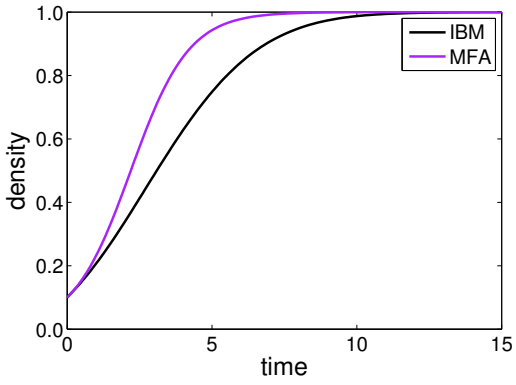
We now compare our derived mean-field models to our averaged discrete results. We begin with the logistic case, equation (1.3.4), observing that there are situations when the MFA provides a good approximation to the averaged discrete behaviour (Figure 1.3(a)). We note that in this region of parameter space, where $P_p/P_m \ll 1$, cells tend not to cluster very much as the rate of movement is much higher than that of proliferation (Figure 1.3(c)). Thus the assumption that lattice site occupancies are independent is valid, and the MFA is a suitable model. However, for many parameter sets, the MFA does not provide a good approximation to the averaged discrete behaviour (Figure 1.3(b)). Clustering of agents can no longer be neglected (Figure 1.3(d)), thus assuming lattice site occupancies are independent is no longer valid due to the correlations that build up between the occupancy of lattice sites. In such cases we can no longer use the MFA and it becomes necessary to develop a model which includes information on the 'patchiness' of the system.

In Figure 1.4, we compare the invading fronts observed from averaged discrete simulations to those obtained from the derived mean-field model, equation (1.3.6). We again find that the MFA is in good agreement with the averaged IBM results when P_p/P_m is small. In the case where P_p/P_m is large, we see that the mean-field model predicts a much faster invasion speed than that predicted by the averaged discrete results. We notice that the shape of the front differs in these different regions: for small P_p/P_m we have a shallow front, whereas for large P_p/P_m the front is steeper.

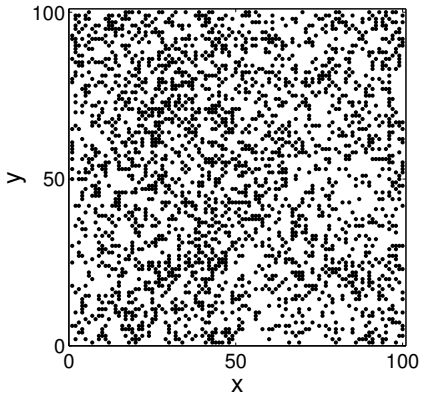
We see that the mean-field models are better able to approximate the averaged discrete simulation results when the motility rates are high relative to those of proliferation. However, when motility rates are low relative to proliferation, the mean-field models do not as accurately approximate the averaged discrete results. This region of parameter space is characterized by the development of clusters from homogeneous initial conditions, or by a sharp invading front for non-homogeneous initial condi-



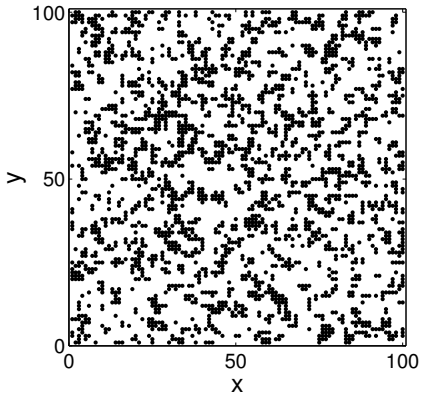
(a) $P_m = 100, P_p = 1$



(b) $P_m = 5 \times 10^{-3}, P_p = 1$



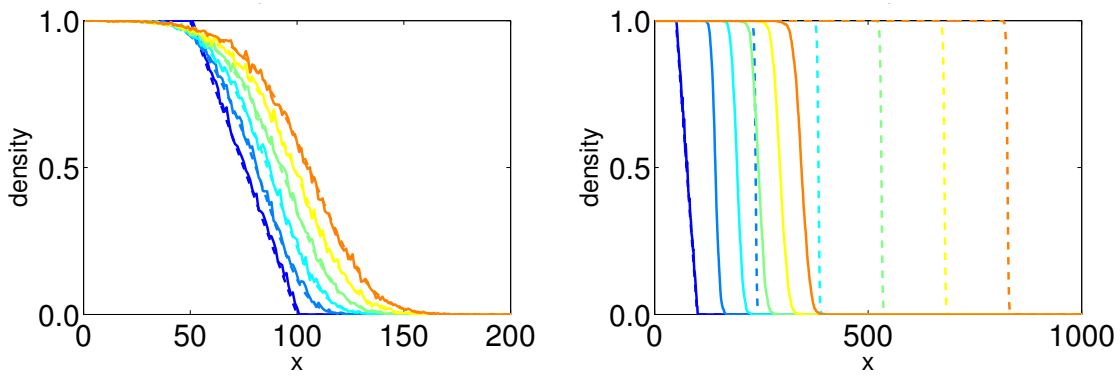
(c) $P_m = 100, P_p = 1$



(d) $P_m = 5 \times 10^{-3}, P_p = 1$

Figure 1.3: (a) We compare the MFA (the logistic model), equation (1.3.4), to the averaged discrete results when the rate of proliferation is low relative to the rate of migration. We see that the MFA provides a good approximation to the averaged discrete behaviour. (b) The rate of proliferation is high relative to the rate of migration in this case, and we see that the MFA no longer provides a good approximation to the averaged discrete behaviour. (c), (d) Snapshots of individual simulations when the lattice is 25% occupied for the parameters used in (a) and (b), respectively. We observe that there is more clustering when the MFA does not provide as good an approximation to the averaged discrete behaviour. The numerical details for production of these plots can be found in Chapter 2.

tions. Many biological systems show high levels of clustering even when they have been seeded randomly, as can be seen in Figure 1.5, thus we must be sure to not use models which depend on the MFA, explicitly or implicitly, in situations where the underlying assumptions may no longer be valid. The accuracy of the MFA can also depend on specific experimental conditions. When bacteria are grown in shaken



(a) $P_m = 1, P_p = 5 \times 10^{-3}$ (b) $P_m = 5 \times 10^{-3}, P_p = 1$

Figure 1.4: (a) The MFA (dashed lines) for an invading front, equation (1.3.6), and the averaged discrete results (solid lines) predict similar population spread when P_p/P_m is relatively small. (b) When P_p/P_m is large, the MFA predicts more rapid population spread than that observed in the averaged discrete case. We also notice that the shape of the population edge varies with the different parameters: it is sharper when P_p/P_m is larger. In both (a) and (b), we have shown the density from $t = 0$ (blue) to $t = 500$ (red) in steps of 100. The details of the relevant numerical methods used in this case can be found in Chapter 3.

liquid cultures, for example, the cells will be well-mixed thus the MFA is an appropriate choice. However, bacteria are frequently grown on agar plates in which case we usually see clumps of bacteria developing from homogeneous initial conditions, implying the MFA will probably no longer be a suitable approximation [Chao and Levin, 1981].

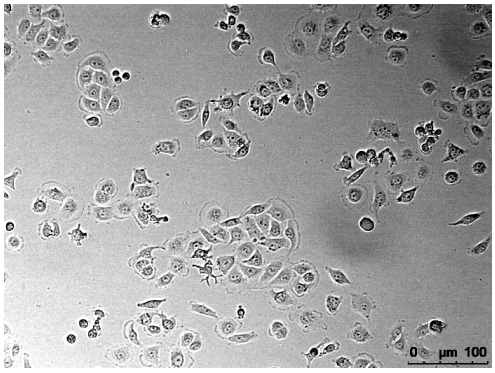


Figure 1.5: Cells are frequently observed to form aggregates over time. The cells in this case are MDA MB231, a breast cancer cell line, and were seeded randomly. Image credit: Matthew Simpson, Queensland University of Technology, Australia.

If we are faced with a biological system in which we cannot assume that the locations of individuals are independent of one another, we should be cautious about using the MFA. We can use IBMs, but these are computationally expensive, therefore we would prefer another option whereby we can include some further information at the level of individual locations without explicitly modelling every agent.

1.3.3 Moment dynamics models

Methods which retain some description of individual behaviour, but without the cost of explicitly modelling each agent, would be ideal for many biological systems. By incorporating additional information into population-level models we can create more sophisticated models, which include some information on the spatial distribution of agents, to approximate the averaged discrete behaviour. In the MFA, we assume that the spatial location of each individual is independent of all others. Of course, this assumption is valid in some situations: when the system is always well-mixed, the location of individuals will be independent of one another. However, this assumption of spatial independence may not always hold, and thus we can improve upon the MFA by relaxing this independence assumption and taking into account the dynamics of pairs of individuals. In the mean-field case, we assume that $\mathbb{P}(\sigma_l, \sigma_m) = \mathbb{P}(\sigma_l)\mathbb{P}(\sigma_m)$. Pair-wise models explictly account for the evolution of $\mathbb{P}(\sigma_l, \sigma_m)$, using master equations in the same way as we showed previously for a single site, $\mathbb{P}(\sigma_l)$. This leads to a larger system of equations to solve than with the MFA, but the system is generally still numerically tractable, and less costly to simulate than an IBM. We discuss the details of deriving pair-wise models in the following chapters. For now, we focus on the development of these models, some of their important contributions, and areas that are currently unexplored.

Moment dynamics models that take into account the dynamics of pairs of individuals will often depend on the triplet behaviour: the evolution of $\mathbb{P}(\sigma_l, \sigma_m)$ depends on

$\mathbb{P}(\sigma_l, \sigma_m, \sigma_n)$. This is similar to the way we saw that the evolution of singlets, $\mathbb{P}(\sigma_l)$, usually depends on pairs, $\mathbb{P}(\sigma_l, \sigma_m)$. We could explicitly account for the dynamics of our triplets [Johnston et al., 2012], which in turn are likely to depend on quadruplets of agents, and so on. Alternatively, we could choose an appropriate closure to express triplet terms using pairs and singlets. This has been investigated in some detail [Dieckmann and Law, 2000; Raghib et al., 2011; Simpson et al., 2014] with the Kirkwood Superposition Approximation (KSA) [Kirkwood, 1935; Kirkwood and Boggs, 1942] being a commonly chosen closure due to its symmetric form and entropy maximizing properties [Singer, 2004].

Whilst moment dynamics models have been used for some time in the physical sciences [Dobrushin, 1986; Grandy, 1987; Mai et al., 1993, 1994; Martin, 1968], it was not until 1992 that they were first applied in a biological setting [Matsuda et al., 1992]. This work demonstrated the improvement on the MFA by using pair and triplet approximations in a Lotka-Volterra model. These models only accounted for nearest-neighbour pairs, assuming that lattice sites were independent beyond this point ($\mathbb{P}(\sigma_l, \sigma_m) \equiv \mathbb{P}(\sigma_l)\mathbb{P}(\sigma_m)$ when l and m are not nearest-neighbours), but paved the way for the use of models such as these in biology, and they have been applied to a wide range of settings since.

Many of the applications of moment dynamics models have continued to focus on ecological systems, for which the spatial distribution of individuals plays a significant role [Ellner, 2001; Kessler and Sander, 2009; Murrell, 2005]. For example, moment dynamics models can help us understand why populations persist in reality when classical models predict they should not [Keeling, 2000], and can be extended to deal with fragmented landscapes [Murrell and Law, 2000]. Advection has also been included, which is an important feature in modelling plankton systems [Young et al., 2001]. Incorporating higher-order dynamics can also be used to explain the reduced speed of invasion observed ecologically [Ellner et al., 1998; Lewis, 2000], as

well as to understand extinction thresholds [Ovaskainen et al., 2002]. Moment dynamics predator-prey models have also been examined in some detail [Barraquand and Murrell, 2013; Murrell, 2005], with the finding that the locations of predators and their prey are inversely correlated [Barraquand and Murrell, 2013], and that increased prey mobility may decrease the final prey density due to increased potential for predator-prey interactions [Murrell, 2005].

Spatial factors also play a key role in plant dynamics, thus this is another important application area in which moment dynamics models have been shown to be successful at reproducing what has been observed in discrete models [Bolker and Pacala, 1997; Takenaka et al., 1997]. The behaviour of plant species frequently depends on dispersal and interaction kernels, which have been successfully included in moment dynamics models [Law et al., 2003; Waddell et al., 2010]. Examining the role of dispersal has shown us that long-range dispersers are more effective at maximizing equilibrium population density [Harada, 1999], unless interspecific competition is strong and then short-range dispersal becomes advantageous [Bolker et al., 2003]. Additionally, the inclusion of high-order moments can mathematically explain the advantages of spatial strategies that ecologists have observed plants adopting in real systems [Bolker and Pacala, 1999].

Another area of mathematical biology in which the spatial distribution of individuals will clearly affect the system behaviour is the spread of infectious diseases [Boots and Sasaki, 1999; Dieckmann and Law, 2000; Keeling, 1999b; Messinger and Ostling, 2009; Sharkey et al., 2006]. Contact between infected and susceptible individuals is necessary for an infection to spread, thus if there is a large amount of clustering, we expect the epidemic to be less severe than the MFA would predict, as fewer infected individuals will be in contact with susceptible individuals. Indeed this has been found to be the case for SIR network models [Keeling, 1999b; Keeling and Eames, 2005], and pair-wise approximations have provided a valid estimation of

the individual behaviour in these cases, with the inclusion of spatial effects leading to lower persistence [Keeling, 1999b; Keeling and Eames, 2005; Webb et al., 2007]. Many earlier moment dynamics models for epidemics account for neighbouring pair-wise interactions only, not considering the longer range behaviour [Filipe and Gibson, 2001; Haraguchi and Sasaki, 2000; Kamo et al., 2007; Parham and Ferguson, 2006]. However, even including just neighbouring pair-wise interactions does still provide a marked improvement when compared to the MFA [Filipe and Gibson, 2001; Haraguchi and Sasaki, 2000], and extensions beyond nearest-neighbour pairs have been considered [Filipe and Maule, 2003]. In addition, stochastic terms have been added, which have been found to lead to different levels of variation when incorporated into a pair-wise, rather than mean-field, model [Dangerfield et al., 2009]. It is even possible, under certain circumstances, to derive exact equations of the infectious time series for certain networks [Sharkey et al., 2013]. To study problems of specific biological relevance, moment dynamics models for epidemics have been extended to account for age-structuring [Keeling et al., 1997] and various vaccination regimes [Keeling, 1999a], as well as asymmetric networks [Sharkey et al., 2006]. Moment dynamics models of epidemics have been increasingly applied to real-world problems. For example, a model of this form was used to analyse the foot-and-mouth epidemic that occurred in Great Britain in 2001 [Ferguson et al., 2001].

Most recently, the importance of spatial structure has been accounted for in cell biology applications [Ascolani et al., 2013; Baker and Simpson, 2010; Simpson and Baker, 2011; Simpson et al., 2013a]. A model for a one-species system undergoing birth-migration-death processes from uniform initial conditions has been developed [Baker and Simpson, 2010]. This model, which uses spatial correlation functions to attempt to correct the mean-field model, shows good agreement with the averaged individual behaviour. Moreover, it has also been used to calibrate experimental data for two different cell types [Simpson et al., 2013a]. For one cell type, where the

rate of migration is sufficiently high relative to that of proliferation, the MFA and correlations model give similar results. However, for the other cell type, where the rate of migration is lower than that of proliferation, the MFA and correlations model give different results. These differences emphasise the importance of using an appropriate model when analysing experimental data, otherwise inacurrate predictions may arise. Of course, many cell-based systems do not begin from spatially uniform initial conditions. For example, tumour invasion and wound healing generally have non-uniform initial conditions. Therefore, moment dynamics models that incorporate non-uniform initial conditions have also been developed, again showing improvement in comparison to the MFA at estimating averaged discrete results [Ascolani et al., 2013; Simpson and Baker, 2011]. Adhesion and contact interactions are important features of many cell-based processes, and correlations are likely to arise from these sorts of interactions; a nearest-neighbour correlations model which includes these phenomena has been developed for systems beginning from non-uniform initial conditions [Ascolani et al., 2013], and a method including longer-range correlations has also been developed [Johnston et al., 2012].

These moment dynamics models have shown a great improvement upon mean-field models across a breadth of applications. However, they are not without their own shortcomings. In the same way that the MFA is limited by its underlying assumptions, moment dynamics models are limited by the higher-order closures that are necessary to produce a tractable model. For instance, pair-wise models do not take into account higher-order structures such as loops, thus are more prone to failure when network connections are strongly localized [Keeling and Eames, 2005; Sharkey, 2011]. Moreover, moment dynamics models can quickly become very complex and unwieldy, even at the level of pairs. They are especially challenging when the system in question begins from non-homogeneous initial conditions [Simpson and Baker, 2011]. Thus developing more tractable models which include pair-wise information is

necessary if these models are to be more widely applicable. Additionally, it is essential that we understand where these models begin to break down, ensuring we restrict their use to feasible applications. In regions where second-order models are no longer viable, we will be required to develop new strategies for deriving continuum models, or rely solely on IBMs.

1.4 Conclusion

In this chapter, we have briefly surveyed some common modelling techniques used in the field of mathematical biology. Phenomenological models, which have been in use for centuries, are simple to implement and generally fairly tractable, allowing us to gain valuable insight into the system under consideration. However, they neglect important behaviour at the level of individuals. For instance, they implicitly assume that the locations of agents are independent of one another (a mean-field assumption). In certain situations, these assumptions are valid and we are able to use mean-field models to accurately represent the biological system in question. However, in biology, we are frequently faced with systems in which the MFA is not appropriate. Thus, we cannot use mean-field models in many of these cases. Instead, we can include the behaviour of every agent by using IBMs, a more recently developed technique. This ensures we do not neglect the interactions of the individuals in our system. However, this improvement comes at a huge cost: not only are IBMs computationally expensive, but we also often lose any potential analysis opportunities.

Most recently, methods have been developed for rigorously deriving the behaviour of a population directly from the individuals of which it is comprised. The simplest of these methods rely on the MFA, thus are only applicable to systems where the spatial locations of agents are independent and, in some cases, are identical to our phenomenological models. We can take one step further than the MFA by examin-

ing the behaviour of pairs of individuals. These higher-order models improve upon the mean-field models, whilst not being as computationally intensive as IBMs that simulate the systems in full.

Of course, there are other modelling methodologies besides those which we have discussed here. One technique that is sometimes used to include some individual information whilst still remaining computationally feasible is hybrid modelling [Gatenby et al., 2007; Kim et al., 2007; Smallbone et al., 2007]. In this class of models, some of the system is represented with continuum models, and only parts of the system which depend strongly on the behaviour at the individual level use an individual modelling approach. Whilst these models may sometimes be more tractable, there are still a number of issues, such as linking the continuum and individual aspects of the model. Similarly to moment dynamics models, other techniques are also borrowed from statistical physics. For example, the one-hole approximation (OHA) uses series expansions to examine a moving cell front [Callaghan et al., 2006] whilst many-body field theory uses a Hamiltonian written in terms of creation and annihilation terms [Dodd and Ferguson, 2009], and other methods use the concept of Fock space which accounts for all the possible states the system could potentially occupy [Birch and Young, 2006]. A technique borrowed from computer science, message passing has been used for modelling epidemics on a network [Karrer and Newman, 2010; Wilkinson and Sharkey, 2014], and it has been shown that pair-based moment-closure models can be seen as a special case of the messaging passing approach [Wilkinson and Sharkey, 2014].

In this thesis, we look at models beyond the mean-field for cellular processes and compare them with mean-field models, demonstrating their improved behaviour at approximating the averaged discrete behaviour. We examine our models across a wide range of parameter space and also discuss situations in which they may not be an appropriate choice. In the next chapter, Chapter 2, we discuss an existing

model [Baker and Simpson, 2010] for including spatial correlations in a one-species birth-death-migration system with homogeneous initial conditions. We reproduce some key results from this work, demonstrating the model's success in a range of parameter space, compared to the MFA. We also highlight some of the difficulties encountered when implementing this model numerically. We then examine a similar system, but with non-homogeneous initial conditions in Chapter 3 (based on the publication Markham *et al.* (2014), Comparing methods for modelling spreading cell fronts, *J. Theor. Biol.*, **353**, 95-103). In this chapter, we compare the MFA, a correlations model [Simpson and Baker, 2011] and an improved version of the OHA [Callaghan et al., 2006] across a range of parameter space and experimental conditions. We find that the MFA and correlations model work best in regions where migration rates are high relative to those of proliferation. In contrast, the OHA is most suitable at approximating the averaged discrete data when migration rates are low relative to those of proliferation. Due to some difficulties in implementing these models, we simplify the model of Baker and Simpson [2010] in Chapter 4 (based on the publication Markham *et al.* (2013), Simplified method for correcting mean-field approximations, *Phys. Rev. E*, **87** 062702). We demonstrate that this simplified model provides a good approximation to the averaged discrete behaviour. By investigating a wider range of parameter space than in Chapter 2, we determine regions where the correlations model is no longer appropriate. This simplified model lends itself more readily to extensions, thus allowing us to investigate a multi-species system in Chapter 5 (based on the publication Markham *et al.* (2014), Incorporating spatial correlations into multispecies mean-field models, *Phys. Rev. E*, **88**, 052713). We show that our multi-species correlations model provides a good approximation to the averaged discrete behaviour in a wide range of parameter space. This is in contrast to the MFA, which is frequently unable to accurately predict the averaged discrete results. We then, in Chapter 6, examine three methods for determining which model may be best suited

to a particular set of experimental data. We find that all the clustering measures we discuss are time dependent, thus will require experimental data collected at a series of times rather than at a single point. Additionally, we find that the best indications of model suitability are those which account for highly localized clustering. Finally, we conclude this thesis in Chapter 7, where we discuss our results and some potential future directions that could be taken.

Chapter 2

A single-species correlations model

In Chapter 1, we described a number of modelling techniques commonly used in mathematical biology. Phenomenological population models, which frequently rely on mean-field assumptions, have played an essential role since the field of mathematical biology was established. However, they do not explicitly account for behaviour at the individual level, which can lead to inaccurate model predictions in some cases. In contrast, IBMs include the behaviour of every individual in the model, at great cost. By incorporating additional spatial information, such as the evolution of pairs of agents, we can retain some of the individual behaviour in population models without the computational cost of simulating every agent. This allows us to more accurately represent the biological system in question than we would be able to do with a mean-field model, but with a model that is less computationally intensive and more amenable to analysis than an IBM. In this chapter we examine an existing model [Baker and Simpson, 2010] which incorporates spatial correlations into a mean-field model for a birth-death-migration process on a regular lattice. We reproduce some of the key results, and discuss some challenges encountered when using this model.

2.1 Introduction

Mean-field models are only suitable under certain conditions. Having models that can be used over a greater range of conditions than those for which the MFA is valid will have invaluable applications in the biological sciences. For instance, mean-field models are often used to estimate parameters from biological data, and the parameters are frequently employed in further modelling studies [Tremel et al., 2009]. However, if the system being examined has significant spatial correlations, there is the risk that the parameters will be inaccurately estimated when using a mean-field approach [Simpson et al., 2013a]. Thus, for cases where we can no longer make standard mean-field assumptions, it is necessary to use models that go beyond the mean-field if we are to have confidence in our predictions. As we have discussed in the previous chapter, we can improve upon mean-field models by incorporating higher-order information, such as the evolution of pair-wise densities. In this chapter, we discuss an existing spatial correlations model from the literature for a one-species system containing agents undergoing proliferation, movement and death on a regular d-dimensional square lattice ($d = [1, 2, 3]$) with volume exclusion [Baker and Simpson, 2010]. We choose to focus our attention on birth-death-migration processes as these cellular events play a vital role in the development and repair of organisms [Gilbert, 2006], as well as determining the survival success of cell populations [Barrandon and Green, 1987]. We begin, in Section 2.2, by describing the system under investigation. We then turn to the different modelling techniques used to represent this system: an IBM (Section 2.3), the MFA (Section 2.4) and a correlations model (Section 2.5). In Section 2.6, we discuss some of the computational complexities of implementing the correlations model, particularly in higher dimensions. Finally, in Section 2.7, we compare results from the correlations model and the MFA to those obtained from averaged discrete simulations, demonstrating the improved performance achieved by including spatial correlations when approximating the averaged discrete density results.

2.2 The system

We consider a one-species contact process with diffusion [Harris, 1974; Liggett, 1999] whereby we have a d-dimensional system on a regular, square lattice with lattice spacing Δ. Initially, agents are placed randomly on the lattice with each site having an occupancy probability of P_i, thereby giving, on average, homogeneous initial conditions. We regard sites as being occupied (state A) or unoccupied (state 0), with the lattice state of site l being described by the variable σ_l. The normalised average agent density is denoted by $c_A(t)$:

$$c_A(t) = \frac{1}{N_{tot}} \sum_i \mathbb{1}_A\{\sigma_i\}, \tag{2.2.1}$$

where N_{tot} is the total number of lattice sites, and $\mathbb{1}_A$ is the indicator function. The agents move to neighbouring lattice sites with a rate P_m per unit time; they proliferate and place a daughter on a neighbouring lattice site with a rate P_p per unit time; and die with a rate P_d per unit time. Proliferation and movement events will only occur if the target site is unoccupied, thereby giving rise to volume exclusion. The target site for movement and proliferation events is chosen at random from the nearest-neighbour sites according to a von Neumann neighbourhood. In Figure 2.1, we see the possible events for an agent with no neighbours in a 1D system, and for an agent with one neighbour in a 1D system.

2.3 Discrete simulations

To simulate sample paths of the system, we use a modified Gillespie algorithm [Baker and Simpson, 2010; Gillespie, 1977], the algorithm for this being as follows:

1. set $t = 0$;

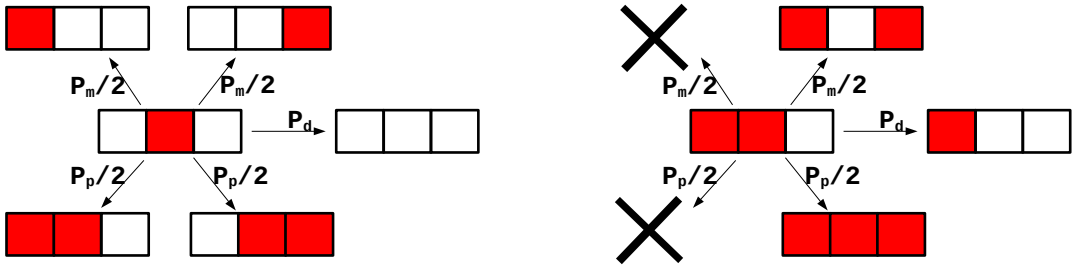


Figure 2.1: An example of the possible events for an agent with no neighbours (left) and with one neighbour (right) in 1D. When an agent has a neighbour, some of the attempted events are not possible, thus are aborted. This can be seen on the right, where the site in the centre cannot move or proliferate into the site on its left as it is already occupied. Occupied sites are red, whilst unoccupied sites are white.

2. initialize the lattice by generating a uniform random number, r_i, in the interval [0,1], for each lattice site. If $r_i \leq P_i$, an agent is placed at that lattice site;

3. with $Q(t)$ being the total number of agents on the lattice at time t, calculate the total propensity function, $a_0 = (P_m + P_p + P_d)Q(t)$;

4. calculate the time, τ, to the next event as $\tau = (1/a_0)\log(1/r_1)$, where r_1 is a uniform random number in the interval [0,1];

5. decide which event occurs by calculating $R = a_0 r_2$, where r_2 is another uniform random number in the interval [0,1]. R is used to deduce which event occurs according to the following:

   - if $R \in [0, P_m Q(t))$, a movement event will be attempted. An agent is chosen uniformly at random, and one of its neighbours in a von Neumann neighbourhood is also chosen uniformly at random as the target site for a movement event. If the target site is empty, the chosen agent moves to that site, otherwise the event is aborted;
   - if $R \in [P_m Q(t), (P_m + P_p)Q(t))$, a proliferation event will be attempted. An agent is chosen uniformly at random, and one of its neighbours in a von Neumann neighbourhood is also chosen uniformly at random as the

target site for a proliferation event. If the target site is empty, a daughter cell is placed in it, otherwise the event is aborted;

- otherwise, a death event occurs. An agent is chosen uniformly at random and removed from the system;

6. update $Q(t)$ depending on which event, if any, occurred;

7. update time by setting $t \to t + \tau$;

8. repeat from step 3 until the stipulated final time is exceeded.

We implement this algorithm, with periodic boundary conditions, in C++ on a $10,000$ lattice in 1D, 100×100 lattice in 2D, and a $20 \times 20 \times 20$ lattice in 3D. These dimensions maintain roughly the same number of sites in total, and are small enough to be computationally feasible. We see the results of a single 2D simulation in Figure 2.2, which shows the development of patches of agents over time, for the specific parameters used in this case. To obtain the evolution of the average cell density, we perform one hundred realisations.

We now derive two methods for approximating the evolution of the averaged cell density for this discrete system. The first is the MFA, and the other is a second-order method which takes into account information about the spatial correlations between agent locations.

2.4 The mean-field approximation

We derive our continuum models using k-point distribution functions, $\rho^{(k)}$, which describe the probability that k-tuplets of sites have given occupancies. For example, $\rho^{(1)}(\sigma_l)$ is the probability that site l is in state σ_l, whilst $\rho^{(2)}(\sigma_l, \sigma_m)$ is the joint probability that sites l and m are in states σ_l and σ_m, respectively. We begin with

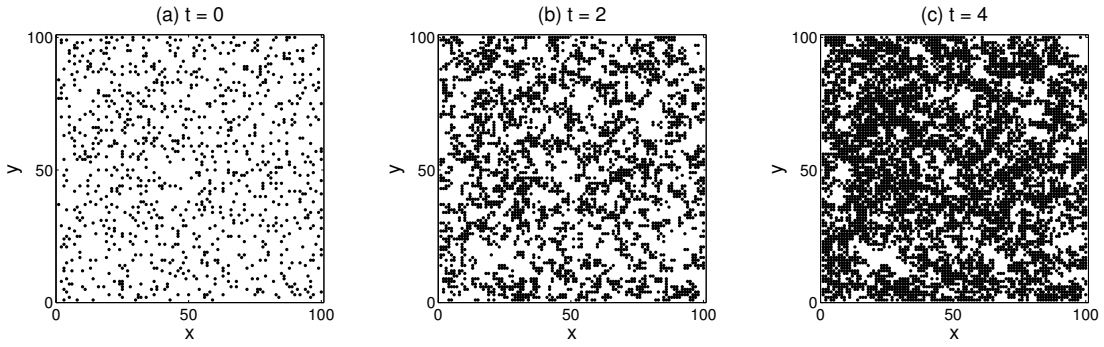


Figure 2.2: We see that, beginning from spatially uniform initial conditions (a), the agents become patchy over time (b)-(c), thus indicating that spatial correlations are important for this set of parameters. Results here are generated using the Gillespie algorithm in Section 2.3 on a 100×100 lattice for $P_m = 0.5$, $P_p = 1.0$, $P_d = 0.1$ and $P_i = 0.1$.

the derivation of the evolution over time of the 1-point distribution functions:

$$\begin{aligned}\frac{\mathrm{d}\rho^{(1)}(A_l)}{\mathrm{d}t} = & \underbrace{P_m \sum_n \frac{\alpha_{n,l}}{z}[\rho^{(2)}(0_l, A_n) - \rho^{(2)}(A_l, 0_n)]}_{\text{Movement to/from neighbouring sites}} \\ & + \underbrace{P_p \sum_n \frac{\alpha_{n,l}}{z}\rho^{(2)}(0_l, A_n)}_{\text{Proliferation from neighbouring sites}} - \underbrace{P_d \rho^{(1)}(A_l)}_{\text{Cell death}},\end{aligned} \tag{2.4.1}$$

where $z = 2d$ (d is the number of spatial dimensions of the system), and $\alpha_{n,l}$ is unity if n and l are nearest neighbours, otherwise it is zero. A schematic of the evolution of $\rho^{(1)}(A_l)$ can be seen in Figure 2.3. In the MFA, we assume that lattice site occupancies are independent of one another. Thus we assume that each 2-point distribution function is the product of its respective 1-point distribution functions: $\rho^{(2)}(\sigma_l, \sigma_m) \equiv \rho^{(1)}(\sigma_l)\rho^{(1)}(\sigma_m)$. Therefore, equation (2.4.1) becomes

$$\begin{aligned}\frac{\mathrm{d}\rho^{(1)}(A_l)}{\mathrm{d}t} = & \underbrace{P_m \sum_n \frac{\alpha_{n,l}}{z}[\rho^{(1)}(0_l)\rho^{(1)}(A_n) - \rho^{(1)}(A_l)\rho^{(1)}(0_n)]}_{\text{Movement to/from neighbouring sites}} \\ & + \underbrace{P_p \sum_n \frac{\alpha_{n,l}}{z}\rho^{(1)}(0_l)\rho^{(1)}(A_n)}_{\text{Proliferation from neighbouring sites}} - \underbrace{P_d \rho^{(1)}(A_l)}_{\text{Cell death}}.\end{aligned} \tag{2.4.2}$$

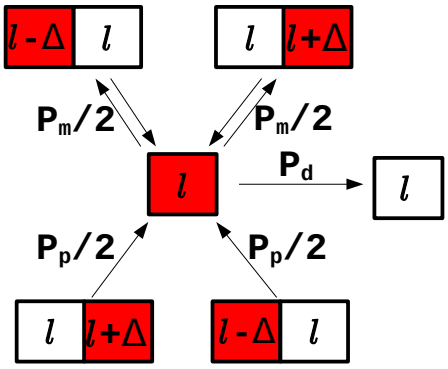


Figure 2.3: A schematic describing the evolution of a 1-point distribution function. The occupancy of site l is determined by the occupancies of neighbouring sites, thereby will depend on 2-point distribution functions. Occupied sites are red, whilst unoccupied are white.

The 1-point distribution function, $\rho^{(1)}(A_l)$, corresponds to the expected density of site l at time t. As our initial conditions are homogeneous, we assume that all $\rho^{(1)}(A_i)$ are identical, with their value being equivalent to the normalised average agent density, c_A. Thus equation (2.4.2) becomes

$$\frac{\mathrm{d}c_A}{\mathrm{d}t} = P_p c_A(1 - c_A) - P_d c_A, \tag{2.4.3}$$

which is simply the (unscaled) logistic equation. The logistic model makes the assumption that systems remain well-mixed thus lattice site occupancies are expected to be independent, and this is the MFA we intuitively expect for the system we have described. We rescale time and density to facilitate comparison of different proliferation and death rates. We rescale density by

$$\bar{c}_A = \left(\frac{P_p}{P_p - P_d}\right) c_A, \tag{2.4.4}$$

giving

$$\frac{\mathrm{d}\bar{c}_A}{\mathrm{d}t} = r\bar{c}_A(1 - \bar{c}_A), \tag{2.4.5}$$

where $r = P_p - P_d$. This can be solved exactly to give

$$\bar{c}_A = \frac{\bar{c}_A(0)e^{rt}}{1 + \bar{c}_A(0)(e^{rt} - 1)}, \tag{2.4.6}$$

where $\bar{c}_A(0)$ is the rescaled initial density. Rescaling time by

$$\bar{t} = (P_p - P_d)t, \tag{2.4.7}$$

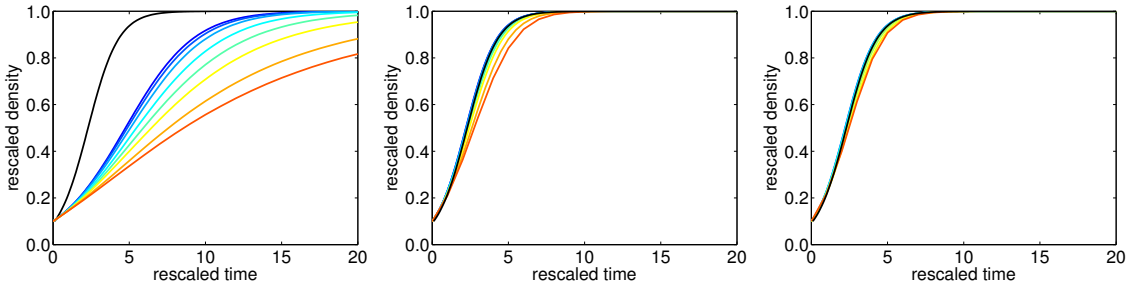
gives us the solution

$$\bar{c}_A(\bar{t}) = \frac{\bar{c}_A(0)e^{\bar{t}}}{1 + \bar{c}_A(0)(e^{\bar{t}} - 1)}. \tag{2.4.8}$$

2.4.1 Comparing the MFA and the IBM

We now compare our MFA and IBM results in Figure 2.4. We examine the performance of our MFA at approximating the IBM for a range of P_p values, keeping $P_m = 1$ and $P_d = 0$. Due to the rescaling of time, equation (2.4.7), the MFA is identical for all values of P_p. However, the averaged discrete results vary with P_p: larger values of P_p lead to slower evolution of the density. Thus, as P_p is increased, the difference between the MFA and averaged discrete results also increases. We see that the disagreement increases as the number of dimensions in our system decreases. This is because the agents are able to move more freely around clustered patches in higher dimensions.

The MFA clearly does not provide a suitable approximation under all circumstances. We now derive a correlations model for this system to determine if including information on the spatial locations of pairs of agents, thereby relaxing the independence assumptions, provides a significant improvement when approximating the averaged discrete behaviour.



(a) 1D (b) 2D (c) 3D

Figure 2.4: Increasing P_p for $P_m = 1$ and $P_d = 0$ gives the same rescaled density if we use the MFA (black), equation (2.4.8). However the averaged simulated densities, using the algorithm in Section 2.3, evolve more slowly as P_p is increased. The values of P_p are $0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5$ and 1, with the lowest value in blue and the highest value in red. We note that the disagreement lessens as the number of dimensions is increased. This is because the agents can more easily circumvent clusters in higher dimensions.

2.5 Deriving the correlations ODEs

The previous section demonstrates that the MFA does not always provide a valid approximation to the averaged discrete behaviour. Thus, we relax the assumption that lattice site occupancies are independent. This means we can no longer assume that the 2-point distribution functions can be expressed as a product of the relevant 1-point distribution functions. We introduce the concept of correlation functions [Baker and Simpson, 2010; Mai et al., 1993, 1994], which relate the 2-point and 1-point distribution functions, providing a measure of the spatial dependence of lattice site occupancy. Assuming translational invariance, we define the correlation functions at a radial distance of $|\mathbf{l} - \mathbf{m}|$ as

$$F_{\lambda,\mu}(|\mathbf{l} - \mathbf{m}|) := \frac{\rho^{(2)}(\sigma_l, \sigma_m)}{\rho^{(1)}(\sigma_l)\rho^{(1)}(\sigma_m)} = \frac{\rho^{(2)}(\sigma_l, \sigma_m)}{c_\lambda c_\mu}, \tag{2.5.1}$$

where λ and μ denote the states of sites $\mathbf{l}$ and $\mathbf{m}$, respectively. The correlation functions depend only on the distance between the lattice sites, as we have assumed an inheritance of the translational invariance and isotropy associated with the initial

conditions. For independent lattice sites, the correlation function will be unity, as $\rho^{(2)}(\sigma_l, \sigma_m) = c_\lambda c_\mu$ in this case. If lattice sites are positively correlated the correlation function will be greater than unity, whilst a correlation function below unity implies the sites are inversely correlated.

We can relate the different correlation functions to each other using conservation equations. For example, we can write the following conservation equation:

$$\rho^{(2)}(A_l, A_m) + \rho^{(2)}(A_l, 0_m) = \rho^{(1)}(A_l), \quad \forall m \neq l. \tag{2.5.2}$$

Using the definition of a correlation function, equation (2.5.1), and that the averaged density of each site is the same for all sites, we have

$$c_A^2 F_{A,A}(|\mathbf{l} - \mathbf{m}|) + c_A c_0 F_{A,0}(|\mathbf{l} - \mathbf{m}|) = c_A, \tag{2.5.3}$$

and thus we can express $F_{A,0}$ in terms of $F_{A,A}$:

$$F_{A,0}(|\mathbf{l} - \mathbf{m}|) = \frac{1 - c_A F_{A,A}(|\mathbf{l} - \mathbf{m}|)}{1 - c_A}. \tag{2.5.4}$$

Similarly, we can express $F_{0,0}$ in terms of $F_{A,A}$, thus the only correlation function we need to consider for the one-species case is $F_{A,A}$. Returning to the derivation of our correlations model, we simplify equation (2.4.1) using the correlation functions and the assumption of translational invariance and isotropy. This gives us

$$\frac{\mathrm{d}c_A}{\mathrm{d}t} = P_p c_A[1 - F_{A,A}(\Delta)c_A] - P_d c_A. \tag{2.5.5}$$

We note that equation (2.5.5) resembles equation (2.4.3), our MFA, however the proliferative term is now dependent on nearest-neighbour correlations ($F_{A,A}(\Delta)$). The death term still does not contain any correlation functions. This is what we expect for our chosen system, as proliferation is dependent on nearest-neighbour site occupancies

whereas death events are independent of the occupancies of neighbouring sites. If lattice site occupancies are independent, $F_{A,A}(\Delta) = 1$, then equation (2.5.5) reduces to equation (2.4.3), as expected. If site occupancies are correlated with one another, we expect $F_{A,A}(\Delta) > 1$, thus slowing down the evolution of c_A. This agrees with the results for the discrete case in Figure 2.4.

To determine the evolution of the correlation functions, we turn to the 2-point distribution functions. A schematic of how $\rho^{(2)}(A_l, A_m)$ evolves in a 1D system is given in Figure 2.5, and the evolution expression is

$$\begin{aligned}\frac{\mathrm{d}\rho^{(2)}(A_l, A_m)}{\mathrm{d}t} = &\underbrace{P_m \sum_{n\neq l} \frac{\alpha_{n,m}}{z}[\rho^{(3)}(A_l, 0_m, A_n) - \rho^{(3)}(A_l, A_m, 0_n)]}_{\text{Movement in and out of site } m} \\ &+ \underbrace{P_m \sum_{n\neq m} \frac{\alpha_{n,l}}{z}[\rho^{(3)}(0_l, A_m, A_n) - \rho^{(3)}(A_l, A_m, 0_n)]}_{\text{Movement in and out of site } l} \\ &+ \underbrace{P_p \left[\sum_{n\neq l} \frac{\alpha_{n,m}}{z}\rho^{(3)}(A_l, 0_m, A_n) + \sum_{n\neq m} \frac{\alpha_{n,l}}{z}\rho^{(3)}(0_l, A_m, A_n)\right]}_{\text{Proliferation from site } n} \\ &+ \underbrace{P_p \frac{\alpha_{l,m}}{z}\left[\rho^{(2)}(A_l, 0_m) + \rho^{(2)}(0_l, A_m)\right]}_{\text{Proliferation when } l \text{ and } m \text{ are nearest neighbours}} - \underbrace{2P_d\rho^{(2)}(A_l, A_m)}_{\text{death}}.\end{aligned} \tag{2.5.6}$$

We simplify, using the assumption of translational invariance and our conservation equations, as for the $\rho^{(1)}$ case. We use the following two conservation equations:

$$\rho^{(2)}(A_l, A_m) = \rho^{(3)}(A_l, A_m, A_n) + \rho^{(3)}(A_l, A_m, 0_n), \tag{2.5.7}$$

and

$$\rho^{(2)}(A_l, A_n) = \rho^{(3)}(A_l, A_m, A_n) + \rho^{(3)}(A_l, 0_m, A_n), \tag{2.5.8}$$

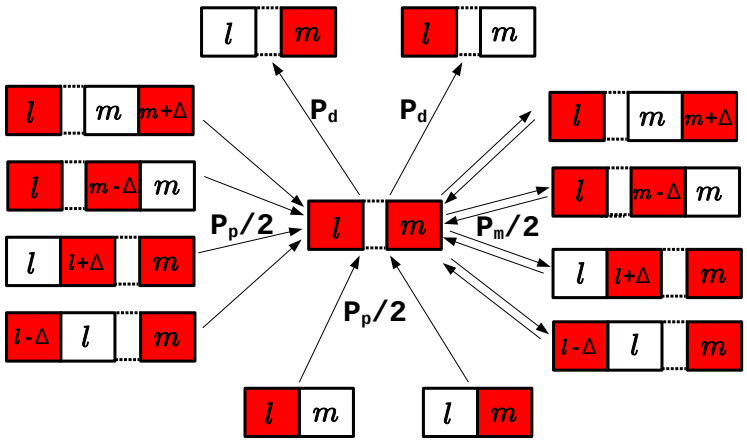


Figure 2.5: A schematic describing the evolution of $\rho^{(2)}(A_l, A_m)$ in 1D shows a dependence on $\rho^{(3)}$ functions. Occupied sites are shown in red whilst unoccupied sites are white. The lattice sites connected by dotted lines are not nearest neighbours.

to write

$$\begin{aligned}\rho^{(3)}(A_l, 0_m, A_n) - \rho^{(3)}(A_l, A_m, 0_n) = &\rho^{(2)}(A_l, A_n) - \rho^{(3)}(A_l, A_m, A_n) \\ &- \rho^{(2)}(A_l, A_m) + \rho^{(3)}(A_l, A_m, A_n),\end{aligned} \tag{2.5.9}$$

which simplifies to

$$\rho^{(3)}(A_l, 0_m, A_n) - \rho^{(3)}(A_l, A_m, 0_n) = \rho^{(2)}(A_l, A_n) - \rho^{(2)}(A_l, A_m). \tag{2.5.10}$$

Thus, we see that the conservation equations can be used to eliminate the $\rho^{(3)}$ terms from our movement expressions in the one-species case, and equation (2.5.6) simplifies to

$$\begin{aligned}\frac{\mathrm{d}\rho^{(2)}(A_l, A_m)}{\mathrm{d}t} =& 2P_m \sum_{n\neq l} \frac{\alpha_{n,m}}{z}[\rho^{(2)}(A_l, A_n) - \rho^{(2)}(A_l, A_m)] \\ &+ 2P_p \sum_{n\neq l} \frac{\alpha_{n,m}}{z}\rho^{(3)}(A_l, 0_m, A_n) \\ &+ 2P_p \frac{\alpha_{l,m}}{z}\left[\rho^{(2)}(A_l, 0_m)\right] - 2P_d\rho^{(2)}(A_l, A_m).\end{aligned} \tag{2.5.11}$$

We notice that the conservation equations have allowed us to eliminate the $\rho^{(3)}$ terms from our movement terms. We can relate the evolution of our correlation functions with that of the $\rho^{(2)}$ functions by using equation (2.5.1) to obtain

$$\begin{aligned}\frac{\mathrm{d}\rho^{(2)}(A_l, A_m)}{\mathrm{d}t} =& c_A^2 \frac{\mathrm{d}F_{A,A}}{\mathrm{d}t}(|\mathbf{l}-\mathbf{m}|) + 2P_p c_A^2[1 - c_A F_{A,A}(\Delta)]F_{A,A}(|\mathbf{l}-\mathbf{m}|) \\ &- 2P_d c_A^2 F_{A,A}(|\mathbf{l}-\mathbf{m}|).\end{aligned} \tag{2.5.12}$$

Using equations (2.5.11) and (2.5.12), we can obtain a description of the evolution of the correlation functions:

$$\begin{aligned}\frac{\mathrm{d}F_{A,A}}{\mathrm{d}t}(|\mathbf{l}-\mathbf{m}|) =& 2P_m \sum_{n\neq l} \frac{\alpha_{n,m}}{z}[F_{A,A}(|\mathbf{l}-\mathbf{n}|) - F_{A,A}(|\mathbf{l}-\mathbf{m}|)] \\ &+ \frac{2P_p}{c_A^2} \sum_{n\neq l} \frac{\alpha_{n,m}}{z}\rho^{(3)}(A_l, 0_m, A_n) \\ &+ \frac{2P_p}{c_A} \frac{\alpha_{l,m}}{z}\left[1 - c_A F_{A,A}(|\mathbf{l}-\mathbf{m}|)\right] \\ &- 2P_p[1 - c_A F_{A,A}(\Delta)]F_{A,A}(|\mathbf{l}-\mathbf{m}|).\end{aligned} \tag{2.5.13}$$

We note that the evolution of $\rho^{(1)}$ depends on $\rho^{(2)}$ terms which, in turn, depend on $\rho^{(3)}$ terms. This trend will generally continue, with the evolution of $\rho^{(k)}$ depending on $\rho^{(k+1)}$ terms, thus we will have a large system of equations to solve. Therefore, we need to close our system at some point for numerical tractability.

2.5.1 System closure

We choose to close our system at $\rho^{(3)}$ for tractability. There are a number of closures at the $\rho^{(3)}$ level in the literature [Dieckmann and Law, 2000; Kirkwood, 1935; Kirkwood and Boggs, 1942; Raghib et al., 2011]. We use the KSA [Kirkwood, 1935; Kirkwood and Boggs, 1942] in our model:

$$\rho^{(3)}(\sigma_l, \sigma_m, \sigma_n) = \frac{\rho^{(2)}(\sigma_l, \sigma_m)\rho^{(2)}(\sigma_l, \sigma_n)\rho^{(2)}(\sigma_m, \sigma_n)}{\rho^{(1)}(\sigma_l)\rho^{(1)}(\sigma_m)\rho^{(1)}(\sigma_n)}. \tag{2.5.14}$$

The KSA is a simple, symmetric expression which makes it widely useable and straightforward to implement. Additionally, the KSA has been shown to maximize the system's entropy [Singer, 2004], further supporting our choice. Use of the KSA can also be justified with a probabilistic argument. If we assume the occupancies of l and n are independent (but the occupancies of l, m and m, n are not necessarily independent), we have

$$\rho^{(2)}(A_l, A_n) = \rho^{(1)}(A_l)\rho^{(1)}(A_n). \tag{2.5.15}$$

Rearranging this we obtain

$$\frac{\rho^{(2)}(A_l, A_n)}{\rho^{(1)}(A_l)\rho^{(1)}(A_n)} = 1. \tag{2.5.16}$$

Using conditional probability arguments, we can write the three-point distribution function as

$$\rho^{(3)}(A_l, 0_m, A_n) = \rho^{(3)}(A_l|0_m, A_n)\rho^{(2)}(0_m, A_n). \tag{2.5.17}$$

As l and n are independent, this becomes

$$\rho^{(3)}(A_l, 0_m, A_n) = \rho^{(2)}(A_l|0_m)\rho^{(2)}(0_m, A_n). \tag{2.5.18}$$

Using conditional probability arguments again, we can rewrite this as

$$\rho^{(3)}(A_l, 0_m, A_n) = \frac{\rho^{(2)}(A_l, 0_m)}{\rho^{(1)}(0_m)}\rho^{(2)}(0_m, A_n). \tag{2.5.19}$$

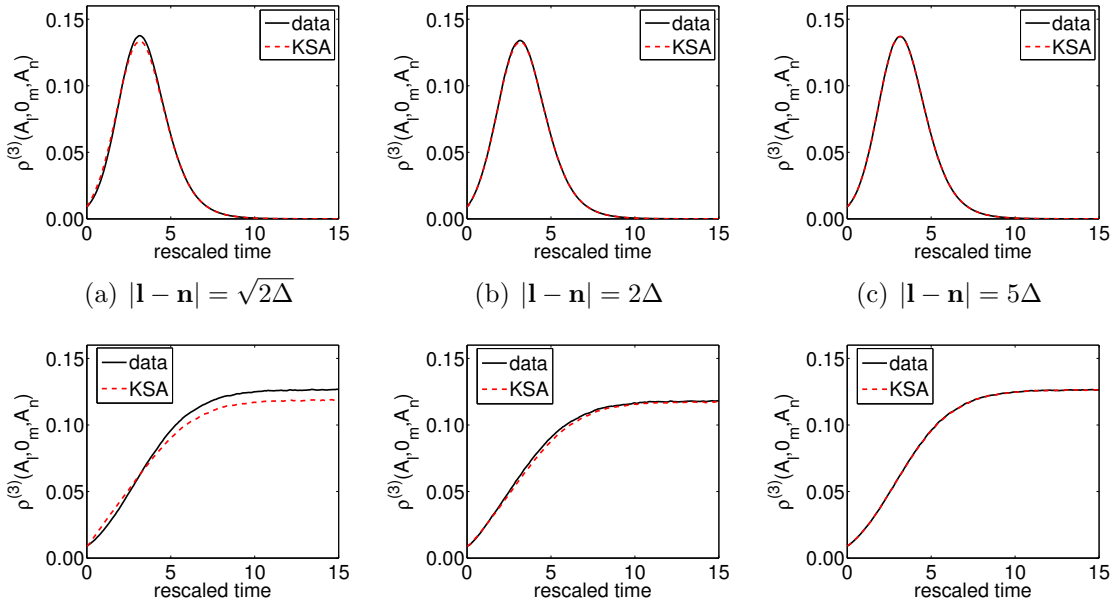
Multiplying the right-hand side of equation (2.5.19) by the expression in equation (2.5.16), we obtain the KSA:

$$\rho^{(3)}(A_l, 0_m, A_n) = \frac{\rho^{(2)}(A_l, 0_m)\rho^{(2)}(A_l, A_n)\rho^{(2)}(0_m, A_n)}{\rho^{(1)}(A_l)\rho^{(1)}(0_m)\rho^{(1)}(A_n)}. \tag{2.5.20}$$

Naturally, being an approximation, the KSA is not always going to perfectly represent the individual behaviour. Using our probabilisitic argument, the KSA is a natural choice of closure when the occupancies of l and n are independent. Of course, the occupancies of l and n are not always going to be independent, particularly when the distance between them is small. Testing our KSA against discrete data (Figure 2.6), we see good agreement between the KSA and discrete data for large $|\mathbf{l}-\mathbf{n}|$, but not such good agreement for smaller values of $|\mathbf{l}-\mathbf{n}|$. We also note that the agreement is worse when the proliferation and death rates are relatively high when compared to migration rates, agreeing with what has been seen by others [Simpson et al., 2014]. Thus we must be cautious using models relying on the KSA in these regions of parameter space.

Having chosen the KSA as our closure, we can obtain a general ODE for the evolution of the correlation functions:

$$\begin{aligned}\frac{\mathrm{d}F}{\mathrm{d}t}(|\mathbf{l}-\mathbf{m}|) = {} & \frac{2P_m}{z}\sum_{n\neq l}\alpha_{n,m}[F(|\mathbf{l}-\mathbf{n}|) - F(|\mathbf{l}-\mathbf{m}|)] \\ & + \frac{2P_p}{1-c_A}\left[1-c_AF(|\mathbf{l}-\mathbf{m}|)\right]\left[1-c_AF(\Delta)\right]\left[\sum_{n\neq l}\frac{\alpha_{n,m}}{z}F(|\mathbf{l}-\mathbf{n}|)\right] \\ & + \frac{2P_p}{c_A}\frac{\alpha_{l,m}}{z}\left(1-c_AF(\Delta)\right) - 2P_p(1-F(\Delta)c_A)F(|\mathbf{l}-\mathbf{m}|),\end{aligned} \tag{2.5.21}$$



(a) $|\mathbf{l}-\mathbf{n}| = \sqrt{2\Delta}$ (b) $|\mathbf{l}-\mathbf{n}| = 2\Delta$ (c) $|\mathbf{l}-\mathbf{n}| = 5\Delta$

(d) $|\mathbf{l}-\mathbf{n}| = \sqrt{2\Delta}$ (e) $|\mathbf{l}-\mathbf{n}| = 2\Delta$ (f) $|\mathbf{l}-\mathbf{n}| = 5\Delta$

Figure 2.6: The KSA, equation (2.5.14), more accurately predicts $\rho^{(3)}$ from the averaged discrete data as the distance between the lattice sites increases. (a)-(c) $P_m = 1, P_p = 0.1, P_d = 0$. (d)-(e) $P_m = 1, P_p = 1, P_d = 0.4$. We see that the agreement is better when proliferation and death are lower relative to movement. In the top row $P_d = 0$ therefore the lattice is eventually completely filled at which point $\rho^{(3)}(A_l, 0_m, A_n)$ is zero.

where F denotes $F_{A,A}$ for simplicity.

2.6 Solving the correlations ODE system

We now discuss solving the system of correlation ODEs, and difficulties encountered therein. In 1D, we have the following equations to solve:

- evolution of the density,

$$\frac{\mathrm{d}c_A}{\mathrm{d}t} = P_p c_A[1 - c_A F(\Delta)] - P_d c_A; \tag{2.6.1}$$

- evolution of the correlation functions for agents a distance Δ apart,

$$\begin{aligned}\frac{\mathrm{d}F(\Delta)}{\mathrm{d}t} =& P_m[F(2\Delta) - F(\Delta)] - 2P_pF(\Delta)[1 - c_AF(\Delta)] \\ &+ \frac{P_p}{c_A}[1 - c_AF(\Delta)] + \frac{P_p}{1 - c_A}F(2\Delta)[1 - c_AF(\Delta)]^2;\end{aligned} \tag{2.6.2}$$

- evolution of the correlation functions for agents separated by a distance x which is greater than Δ,

$$\begin{aligned}\frac{\mathrm{d}F(x)}{\mathrm{d}t} =& P_m[F(x - \Delta) - 2F(x) + F(x + \Delta)] - 2P_pF(x)[1 - c_AF(\Delta)] \\ &+ \frac{P_p}{1 - c_A}[1 - c_AF(\Delta)][1 - c_AF(x)][F(x - \Delta) + F(x + \Delta)].\end{aligned} \tag{2.6.3}$$

For a finite lattice, we can express the correlation functions up to the maximum lattice distance. However, we expect correlations to be negligible at large distances. Thus, in addition to using moment closure for the third-order terms, we also need to choose a distance at which to truncate our system of ODEs. Beyond this point, $x_{\max}$, we assume lattice sites are uncorrelated, thus $F(x \geq x_{\max}) = 1$. We test our chosen value of $x_{\max}$ by including the next point, $x_{\max} + 1$, in our system of equations and confirming the results are visually identical.

In 2D and 3D, the system of equations becomes increasingly more complex due to the non-uniformity of the distances between lattice sites. For example, in Figure 2.7, we look at the first six lattice distances, and the distances upon which they depend, in 2D. The system of equations in 2D takes the following form:

$$\frac{\mathrm{d}c_A}{\mathrm{d}t} = P_pc_A[1 - c_AF(\Delta)] - P_dc_A,$$

$$\begin{aligned}\frac{\mathrm{d}F(\Delta)}{\mathrm{d}t} =& -2F(\Delta)P_p[1 - F(\Delta)c_A] + \frac{P_m}{2}\left[2F(\sqrt{2}\Delta) + F(2\Delta) - 3F(\Delta)\right] \\ &+ \frac{P_p}{2c_A}[1 - c_AF(\Delta)] + \frac{P_p}{2(1 - c_A)}[1 - c_AF(\Delta)]^2\left[2F(\sqrt{2}\Delta) + F(2\Delta)\right],\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}F(\sqrt{2}\Delta)}{\mathrm{d}t} =& -2F(\sqrt{2}\Delta)P_p[1 - F(\Delta)c_A] + \frac{P_m}{2}\left[2F(\Delta) + 2F(\sqrt{5}\Delta) - 4F(\sqrt{2}\Delta)\right] \\ &+ \frac{P_p}{2(1 - c_A)}[1 - c_AF(\Delta)]\left[1 - c_AF(\sqrt{2}\Delta)\right]\left[2F(\Delta) + 2F(\sqrt{5}\Delta)\right],\end{aligned}$$

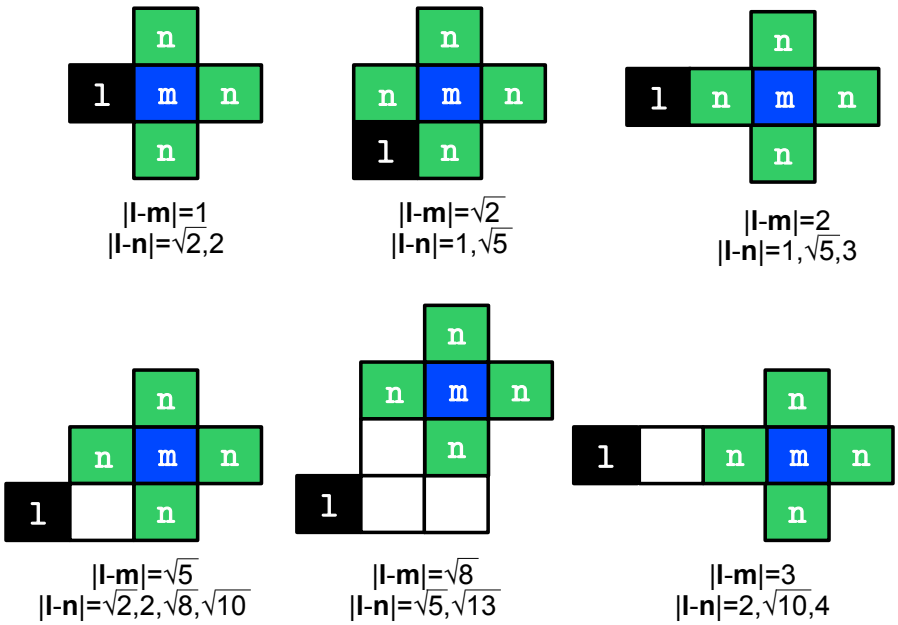


Figure 2.7: In 2D the distances between lattice sites are non-uniform leading to practical difficulties when numerically encoding our equations.

$$\frac{\mathrm{d}F(2\Delta)}{\mathrm{d}t} = -2F(2\Delta)P_p[1-F(\Delta)c_A] + \frac{P_m}{2}\left[F(\Delta)+2F(\sqrt{5}\Delta)+F(3\Delta)-4F(2\Delta)\right] + \frac{P_p}{2(1-c_A)}\left[1-c_AF(\Delta)\right]\left[1-c_AF(2\Delta)\right]\left[F(\Delta)+2F(\sqrt{5}\Delta)+F(3\Delta)\right],$$

$$\begin{aligned}\frac{\mathrm{d}F(\sqrt{5}\Delta)}{\mathrm{d}t} = &-2F(\sqrt{5}\Delta)P_p[1-F(\Delta)c_A] \\ &+\frac{P_m}{2}\left[F(\sqrt{2}\Delta)+F(2\Delta)+F(\sqrt{8}\Delta)+F(\sqrt{10}\Delta)-4F(\sqrt{5}\Delta)\right] \\ &+\frac{P_p}{2(1-c_A)}\left[1-c_AF(\Delta)\right]\left[1-c_AF(\sqrt{5}\Delta)\right] \\ &\times\left[F(\sqrt{2}\Delta)+F(2\Delta)+F(\sqrt{8}\Delta)+F(\sqrt{10}\Delta)\right],\end{aligned}$$

$$\frac{\mathrm{d}F(\sqrt{8}\Delta)}{\mathrm{d}t} = -2F(\sqrt{8}\Delta)P_p[1-F(\Delta)c_A] + \frac{P_m}{2}\left[2F(\sqrt{5}\Delta)+2F(\sqrt{13}\Delta)-4F(\sqrt{8}\Delta)\right] + \frac{P_p}{2(1-c_A)}\left[1-c_AF(\Delta)\right]\left[1-c_AF(\sqrt{8}\Delta)\right]\left[2F(\sqrt{5}\Delta)+2F(\sqrt{13}\Delta)\right],$$

$$\frac{\mathrm{d}F(3\Delta)}{\mathrm{d}t} = -2F(3\Delta)P_p[1-F(\Delta)c_A] + \frac{P_m}{2}\left[F(2\Delta)+2F(\sqrt{10}\Delta)+F(4\Delta)-4F(3\Delta)\right] + \frac{P_p}{2(1-c_A)}\left[1-c_AF(\Delta)\right]\left[1-c_AF(3\Delta)\right]\left[F(2\Delta)+2F(\sqrt{10}\Delta)+F(4\Delta)\right].$$

In the same way as for the 1D case, we need to choose a maximum distance above which all correlation functions are assumed to be unity. As we can see from the equations shown above, adding in further distances is not as straightforward as for the 1D case. This means that if our system has long-range correlations it will be very tedious to implement in practice. We have truncated at a distance of 3Δ in 2D and 3D for the practical reasons we have discussed. We use this distance as the results were found to not be visually different from truncating at $\sqrt{10}\Delta$, for the parameters under consideration.

We solve our system of equations using a fourth-order Runge Kutta routine (RK4) with $\delta t = 0.1$ [Press et al., 2007]. Smaller time steps were tested to ensure this is a suitable choice. Again, we test our code by comparing our results to those obtained from the averaged discrete simulations and MFA in the high movement limit, where all results can be seen to be similar (Figure 2.8(a)). We confirm that our correlation functions are close to unity for high rates of migration (Figure 2.8(b)), as the system is not expected to display significant spatial correlations in these cases. Additionally, we test our code against that used in Baker and Simpson [2010], confirming that the results are identical.

2.7 Results of the correlations ODE system

We now compare the results from our correlations model and MFA to those obtained from average discrete simulations. In Figure 2.8(c), we see that when we are no longer in the high-movement parameter region, the MFA does not as closely predict the averaged discrete results. However, the correlations model provides a fairly good approximation to the averaged discrete results. We see in Figure 2.8(d) that the correlation functions are greater than unity, thus demonstrating that assuming lattice sites are independent (all $F \equiv 1$) is not valid for this set of parameters. In Figure

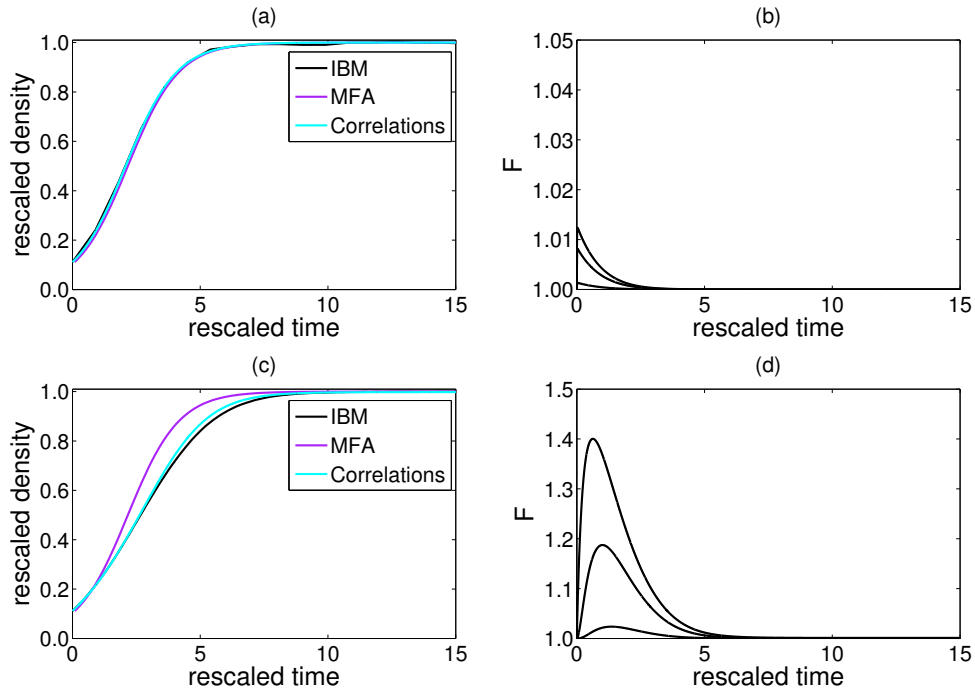


Figure 2.8: (a) When migration rates are high, relative to proliferation, the IBM (using the algorithm in Section 2.3), MFA (equation (2.4.8)), and correlations model (system of equations in Section 2.6) density results are all in good agreement. (b) We see that for high migration rates, the correlation functions remain near to unity. (c) When migration rates are lower, the MFA does not as accurately predict the averaged discrete results, but the correlations model provides a reasonable agreement. (d) For lower migration rates, the correlation functions are noticeably above unity at low times. In both cases, we consider a 2D system. The parameters for (a) and (b) are $P_m = 500, P_p = 1, P_d = 0.1$. The parameters for (c) and (d) are $P_m = 1, P_p = 1, P_d = 0.1$. (b), (d) Correlation functions shown for distances of Δ (highest curve), 2Δ and 3Δ (lowest curve).

2.9, we see the improved performance of the correlations ODE system, compared to the MFA, at predicting the averaged discrete density of the system over time when $P_d = 0$. We note that as P_p is increased, relative to P_m, the deviation between the MFA and IBM results increases. However, the correlations model provides an improvement upon the MFA, yielding a much better prediction of the averaged discrete behaviour. The steady-state behaviour is accurately predicted by both the MFA and correlations model, thus it is only in the transient regions that there is any disagreement. However, in Figure 2.10, we see that including death in our system leads to a

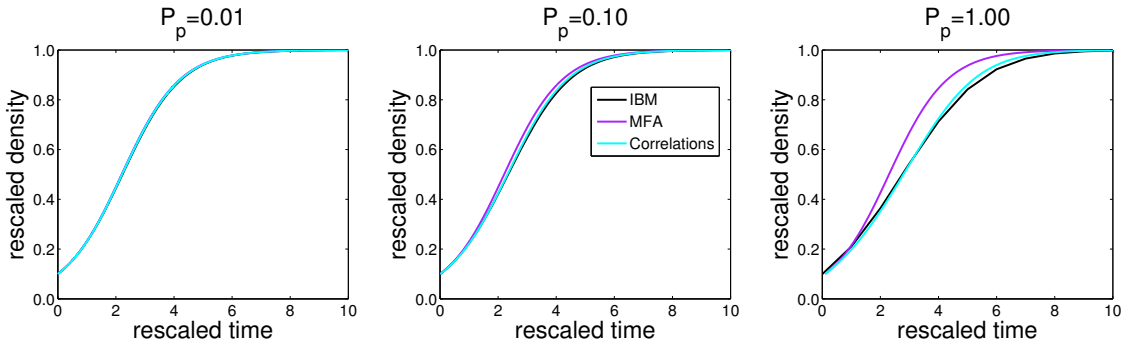


Figure 2.9: As the rate of proliferation increases, relative to that of migration, we see that the MFA, equation (2.4.8), can no longer accurately predict the averaged discrete behaviour, using the algorithm in Section 2.3. In contrast, our correlations model (using the equations in Section 2.6) provides good agreement with the averaged discrete results. In these cases, $P_m = 1, P_d = 0$ and we are in 2D. We see that the MFA accurately predicts the asymptotic behaviour when there is no death in our system, but there are discrepancies in the transient results.

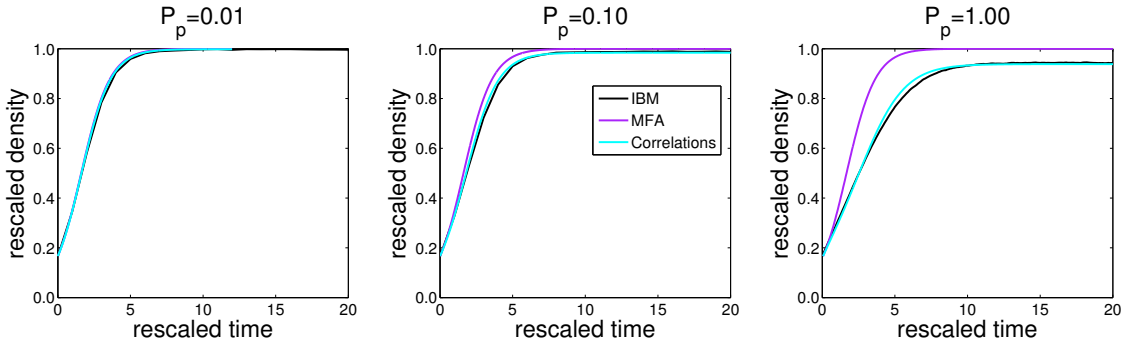


Figure 2.10: We see that with non-zero death rates the MFA, equation (2.4.8), does not always accurately predict the asymptotic or transient behaviour, whereas the correlations model (using the equations in Section 2.6) provides good agreement with the averaged individual results. In all cases, $P_m = 1, P_d = 0.4P_p$ and we are in 2D. We again notice that as the rate of proliferation increases relative to that of migration, the agreement between the MFA and IBM worsens.

different prediction for the asymptotic density from the MFA than from the IBM or correlations model. We once again observe that the disagreement between the MFA and averaged discrete results increases as the proliferation rate increases. Thus, for some regions of parameter space, we cannot rely on the MFA to predict the transient or asymptotic density obtained by our IBM. However, our correlations model provides a good approximation to the averaged discrete behaviour, both transiently and

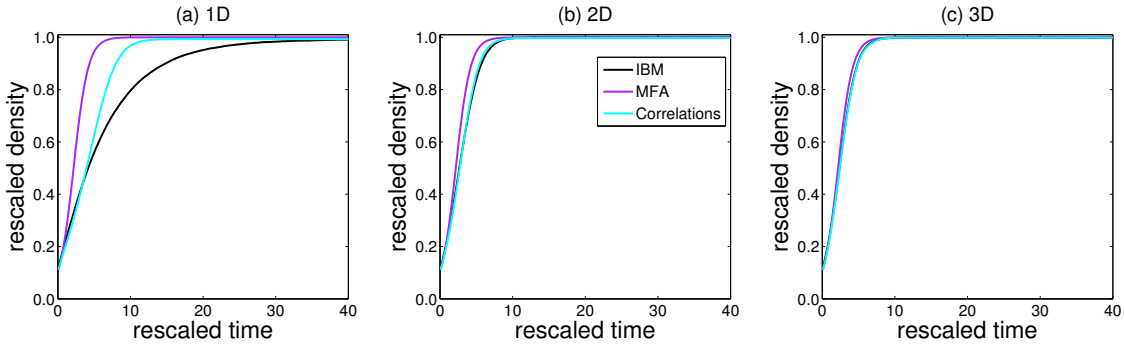


Figure 2.11: Comparing the models in (a) 1D, (b) 2D and (c) 3D for $P_m = 1.0, P_p = 1.0, P_d = 0.1$. For the same parameters, the MFA, equation (2.4.8), and the correlations ODE (using the equations in Section 2.6) become better predictors of the averaged discrete behaviour as the number of dimensions increases, with the correlations model always providing an improvement to the MFA.

asymptotically, and thus is a suitable approximation in these regions of parameter space.

We compare results for the 1D, 2D and 3D cases using the same parameters for each dimension in Figure 2.11. We notice that as the number of dimensions increases, the correlations ODE system better approximates the discrete system for the same set of parameters. This is for the same reason that the MFA performs better in higher dimensions, as we saw in Figure 2.4. Namely, that it is easier for agents to circumvent a cluster of other agents in higher dimensions.

We also vary the initial density of our system, P_i, as seen in Figure 2.12. We find that a lower initial density leads to slightly worse transient agreement between the correlations model and the averaged discrete behaviour. This is because our initial conditions are random: when cells are first placed on the lattice, there are no correlations and these then build up as a result of proliferative events. Let us consider two systems: for the first system we set $P_i = P_x$, and for the second system $P_i = 2P_x$. When the first system reaches a density of $2P_x$, there will already be clustering, but the second system will be spatially uncorrelated at this point as this is the starting point. Thus we see that we obtain a greater level of clustering when the initial

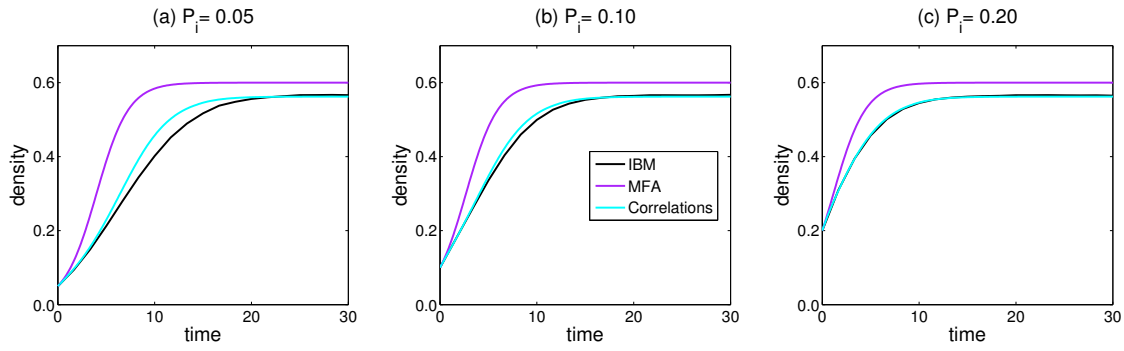


Figure 2.12: Varying the initial density, P_i, in 2D affects the agreement between the continuum models (MFA, equation (2.4.8), and the correlation equations given in Section 2.6) and the averaged discrete behaviour. We see that as the initial density increases, the agreement between the continuum models and the averaged discrete behaviour improves. In this case, we have not rescaled time or density in order to make the changes clearer. The parameters in this case are $P_m = 1, P_p = 1, P_d = 0.4$.

density is lower, thereby causing a decrease in the accuracy of the correlations model at predicting the averaged discrete behaviour. This is also parameter dependent and, for many parameters, the agreement between the correlations model and the averaged discrete results is excellent for a very wide range of initial densities.

Including correlations in our mean-field model provides a great improvement upon the MFA for systems which do not remain well-mixed. Thus it is essential we include correlation effects in many of our biological models. However, including correlations is not completely straightforward to implement in practice as it is tedious to determine and encode the equations for each lattice distance in 2D and 3D. This could be automated, but the number of ODEs is still problematic, and it would be preferable to provide a more comprehensive remedy. The case we have discussed here is the simplest, whereby we have one species beginning from homogeneous initial conditions. For more complex systems, this method would rapidly become intractable, especially if long-range correlations are significant.

2.8 Conclusion

In this chapter, we have demonstrated a method from the literature [Baker and Simpson, 2010] for incorporating spatial correlations into mean-field models for systems with homogeneous initial conditions. This method performs significantly better than the MFA in terms of predicting the averaged discrete behaviour, and thus it is preferable to use this model, particularly if we know our system forms aggregates. The method we have discussed in this chapter is only valid for homogeneous initial conditions: a feature which is not present in all biological systems. In the next chapter, we examine a correlations model, as well as an alternative technique, for modelling spreading cell fronts, beginning from non-uniform initial conditions. We show that the different models provide valid approximations to the averaged discrete space in different areas of parameter space, and discuss which model might be best for a given scenario.

Chapter 3

Comparing methods for modelling spreading cell fronts

In the previous chapter, we discussed a spatial correlations model [Baker and Simpson, 2010] which provides a more accurate approximation of the averaged discrete density results than the MFA. We observed that, in certain regions of parameter space, clusters of agents aggregate over time despite uniform initial conditions, and this leads to the break down of the MFA. The model derivation was based on the assumption that the agents were initially uniformly distributed at random, thus the model could be simplified due to the initial translational invariance of the lattice. What if we begin with non-homogeneous initial conditions and can no longer assume translational invariance, such as is the case for cell invasion problems? In this chapter, we discuss the biological motivation for modelling spreading cell fronts, and some mathematical techniques which can be applied to model systems of this form. We derive the MFA and correlations models for this case [Simpson and Baker, 2011], and examine their success at approximating the discrete behaviour both transiently and asymptotically. Additionally, we derive and extend another method: the one-hole approximation (OHA) [Callaghan et al., 2006], again examining both its transient and

asymptotic behaviour across a range of parameter space. We compare the methods used, discussing their relative merits for different biological situations.

3.1 Introduction

Advancing fronts of cells are frequently observed experimentally [Maini et al., 2004a,b; Simpson et al., 2013b, 2007b]. For example, in Figure 3.1, we see an advancing front of murine fibroblast 3T3 cells from an *in vitro* experiment [Simpson et al., 2013b; Todaro and Green, 1963]. The phenomenon of cell invasion is essential in many physiological processes: embryonic development hinges on the spatial advancement of cells [Simpson et al., 2007b], and wounds could not heal without it [Maini et al., 2004a,b]. Additionally, it is important in tissue engineering [Sengers et al., 2007, 2009], which relies on the ability of fronts of cells to move into unoccupied regions. Less desirably, moving fronts of cells are a major factor in disease progression, most notably in cancer [Allred, 2010; Gatenby and Gawlinski, 1996; Swanson et al., 2003]. An important clinical feature is the sharpness of the front, which is determined by the relative rates of migration and proliferation; a shallow front can lead to difficulties when surgically removing a tumour [Swanson et al., 2003]. Given their importance biologically, it is hardly surprising that moving cell fronts have been the focus of many mathematical modelling studies.

As we discussed in Chapter 1, advancing fronts of cells are classically modelled using the FK equation [Fisher, 1937; Kolmogorov et al., 1937] which, for initial conditions of the form of equation (1.1.5), has a travelling wave solution with constant shape and speed. The asymptotic wave speed (as $t \to \infty$), v_f, for these initial conditions, is $2\sqrt{Dr}$ where D is the diffusivity of the cells, and r their proliferation rate [Murray, 2002]. Measuring the wavespeed experimentally does not allow us to determine unique values for D and r, making additional experimental observations

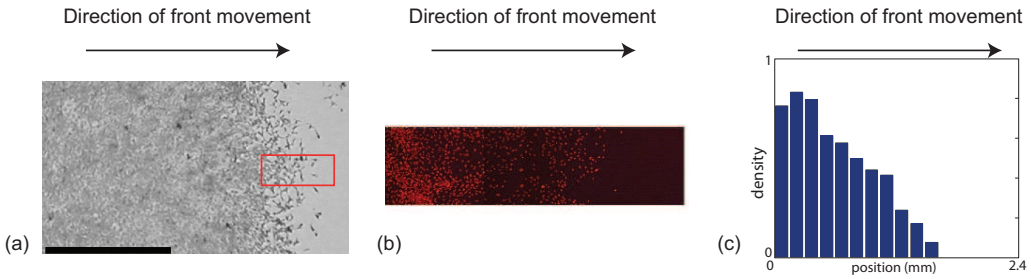


Figure 3.1: Experimentally observed moving front of murine fibroblast 3T3 cells. (a) A snapshot of the cells invading the space to the right. (b) The highlighted region from (a) where the cell nuclei have been stained. (c) The calculated density profile showing the shape of the front. The experimental details for producing images such as these can be found in Simpson et al. [2013b]. Figure credit: Parvathi Haridas, Queensland University of Technology, Australia.

necessary [Sengers et al., 2007; Simpson et al., 2013b]. Moreover, even once the travelling wave has been established, the FK equation, which represents the mean-field behaviour, is not always an accurate representation of the behaviour of a moving front of cells, due to the stochastic nature of the underlying processes [Khain et al., 2011; Lewis, 2000]. Thus, whilst it may be possible to fit experimental data to solutions of the FK model, this does not necessarily lead to accurate parameter estimation; something that is frequently overlooked in models of moving cell fronts [Sengers et al., 2007; Tremel et al., 2009]. This has led to the development of alternative methods for modelling moving cell fronts, some of which we shall now discuss.

Agent-based models, as described in Chapter 1, have been used to examine moving cell fronts in many areas of cell biology [Cheng et al., 2006; Dormann and Deutsch, 2002; Mani et al., 2002]. They are also often used in conjunction with continuum models to provide a multiscale modelling framework [Simpson et al., 2007a]. Discrete models are not confined to any particular region of parameter space, but are limited by their computational cost, and lack of analytical tractability. Thus, whilst they play a vital role in modelling advancing fronts of cells, we would also like to have simpler, more tractable methods approximating the behaviour of moving cell fronts.

When cells proliferate significantly more rapidly than they migrate, we expect a

sharp front [Swanson et al., 2003] with the region behind the front almost completely filled with cells. Under these conditions, we are able to predict the asymptotic front speed using the OHA [Callaghan et al., 2006]. This method uses series expansions to provide a correction term to the front speed for the case without migration, which can be calculated exactly. The OHA agrees well with discrete simulations when proliferation rates are significantly higher than those for migration, and can be extended to deal with more than one hole (unoccupied site) behind the front. However, the method of Callaghan et al. [2006] is only derived for a narrow range of models where a cell either attempts to move or proliferate at every time step, without ever resting. Additionally, experimental results do not always exhibit the asymptotic travelling front behaviour. The following three assays highlight some of the different experiments which can be used to obtain data for travelling fronts.

1. A single moving front is allowed to develop over a long period of time (greater than 100 hours) [Maini et al., 2004a,b]. These experiments are likely to allow for travelling front behaviour to be exhibited. However, they are not as straightforward to carry out as the same experiment over shorter timescales due to difficulties with keeping the cells alive for long periods of time, and maintaining a constant environment.

2. A single moving front is allowed to develop over a short period of time (less than 24 hours). The results of these experiments are not on sufficiently long timescales to produce asymptotic travelling front behaviour [Sengers et al., 2009], but are more feasible experimentally.

3. Two opposingly directed fronts come together. For instance, when a thin strip [Liang et al., 2007; Rodriguez et al., 2005; Valster et al., 2005; Young et al., 2012] or small hole [Young et al., 2012] of cells is removed from a monolayer. In this set-up, the artificially created gap is closed, thus the system may never

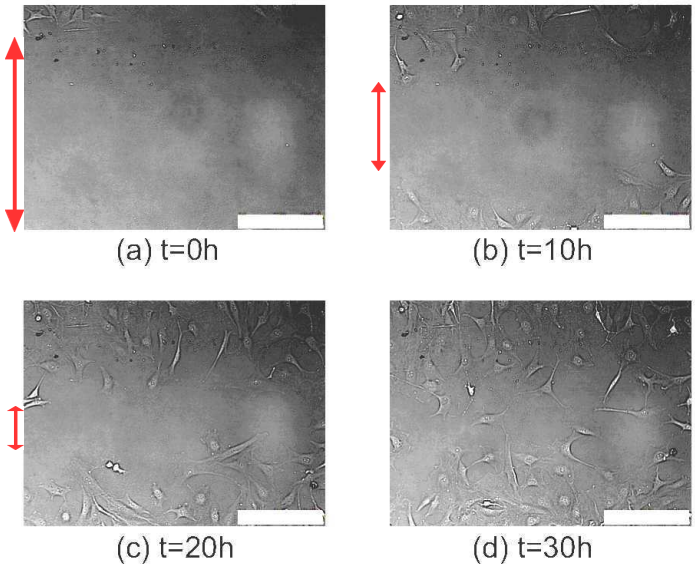


Figure 3.2: Experimental results of a scratch assay with 3T3 cells. We see that no travelling front is established, as the two fronts quickly collide as they approach from either side of the gap. The white bar corresponds to $250\mu m$ and the red arrows give an approximate indication of the width of the scratch at each time point. In (d) there is no arrow as the two fronts have begun to meet in places. Figure credit: Matthew Simpson, Queensland University of Technology, Australia.

reach the asymptotic travelling front speed. For example, the protocol in Liang et al. [2007] allows between eight and eighteen hours for the scratch to close. Given that typical cell doubling times are of a similar magnitude, we do not expect the asymptotic speed to have been reached before the fronts from either side of the scratch become interwoven. We see an example of this in Figure 3.2, where a scratch assay is performed with murine fibroblast 3T3 cells. Within 30 hours, we see the two fronts meeting.

As many experiments follow the second and third methods, it is often important to be able to predict the transient front behaviour as well as the asymptotic speed.

As we discussed in the previous chapters, moment dynamics models incorporate increasingly greater degrees of information into the mean-field model by taking into account the dynamics of cell pairs, triplets, and so forth. Simpson and Baker [2011]

developed a pairwise approximation (PWA), using the KSA [Kirkwood, 1935; Kirkwood and Boggs, 1942] (equation (2.5.14)) as their closure, to describe the behaviour of a moving front. This model is similar to that developed in the previous chapter. However, we are now forced to relax the assumption of translational invariance, which leads to a larger system of equations to solve as we must now keep track of the occupancy of each lattice site. This model allows the transient behaviour of the moving front to be examined and shows improvement on the MFA at approximating the averaged discrete transient behaviour [Simpson and Baker, 2011]. However, the asymptotic behaviour of the model has not been considered.

In this chapter, we examine different methods for modelling spreading cell fronts in detail. We do so over a range of proliferation and migration parameters, comparing our results to those from averaged discrete simulations. Moreover, we discuss which methods are best suited to a given parameter regime and experimental timescale. In Section 3.2, we begin with a description of the discrete Gillespie algorithm used to produce the averaged discrete results. We then discuss the three methods used to approximate the averaged discrete results: the MFA, the PWA, and the OHA. Next, in Section 3.3, we present results focussing on the transient behaviour, evaluating the relative performance of each method. Following this, we turn to the asymptotic results and examine the methods in a range of parameter space. We conclude, in Section 3.4, by discussing the strengths and weaknesses of the methods in question, and which methods might best suit some specific experimental examples.

3.2 The methods

In this section we discuss the methods used to model our system. We consider a 1D, volume-excluding process on a regular lattice with a lattice spacing of $\Delta = 1$. Cells move to neighbouring sites at a rate P_m per unit time and proliferate at a rate P_p per

unit time. Target sites for proliferation and movement events are chosen at random from a von Neumann neighbourhood. If the target site is already occupied, the event is aborted. This is similar to the system discussed in the previous chapter, but we neglect death and the initial conditions are now spatially varying.

3.2.1 The discrete model

For our discrete simulations we implement a Gillespie algorithm in C++ on a 1D lattice with $N = 2000$ sites. We conduct 10,000 individual realisations to estimate the averaged behaviour. Initially, the l^{th} lattice site is occupied with probability

$$C(l) = \begin{cases} 1, & 1 \leq l < x\,, \\ 1 - \left[\dfrac{l-x}{100-x}\right], & x \leq l < 100\,, \\ 0, & 100 \leq l \leq N\,, \end{cases} \tag{3.2.1}$$

where we can alter the steepness of our initial ramp by varying x. We use a Gillespie algorithm to update our system as in the previous chapter, but now with spatially varying initial conditions, which alters the second step. The algorithm is as follows:

1. set $t = 0$;

2. initialize the lattice by generating a uniform random number, r_{il}, in the interval [0,1], for each lattice site. If $r_l \leq C(l)$, an agent is placed at that lattice site;

3. with $Q(t)$ being the total number of agents on the lattice, calculate the total propensity function, $a_0 = (P_m + P_p)Q(t)$;

4. calculate the time, τ, to the next event using $\tau = (1/a_0)\log(1/r_1)$, where r_1 is a uniform random number in the interval [0,1];

5. decide which event occurs by calculating $R = a_0 r_2$, where r_2 is another uniform random number in the interval [0,1]. R is used to deduce which event occurs according to the following:

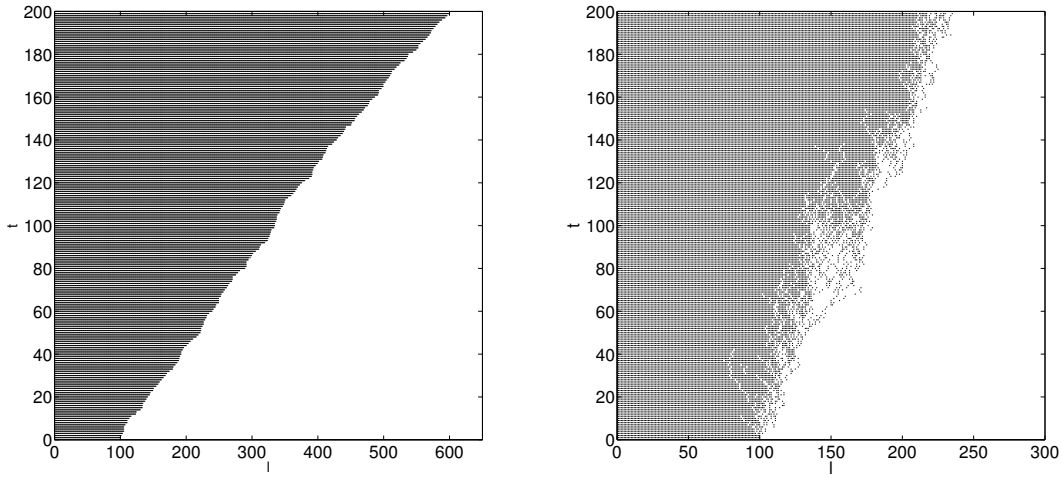
   - if $R \in [0, P_m Q(t))$, a movement event will occur. An agent is chosen uniformly at random, and one of its neighbours is also chosen uniformly at random as the target site for a movement event. If the target site is empty, the chosen agent moves to that site, otherwise the event is aborted;
   - if $R \in [P_m Q(t), (P_m + P_p)Q(t)]$, a proliferation event will occur. An agent is chosen uniformly at random, and one of its neighbours is also chosen uniformly at random as the target site for a proliferation event. If the target site is empty, a daughter cell is placed in it, otherwise the event is aborted;

6. update $Q(t)$ depending on which event, if any, occurred;

7. update time by setting $t \to t + \tau$;

8. repeat from step 3 until the stipulated final time is reached.

We apply reflecting boundary conditions at $l = 1$ and $l = N$. To determine the front speed, v, we track the location where the averaged cell density is 0.5, and calculate the average velocity from this information. Once v is no longer changing with time, we have reached the asymptotic travelling front speed, v_f. As before, we test our discrete code by comparing it to the MFA in the high movement limit, confirming that the results are visually identical. We present two examples of single discrete simulations in Figure 3.3. In Figure 3.3(a), P_p/P_m is large and we see a sharp front, with the majority of sites behind the front being occupied. In Figure 3.3(b), P_p/P_m is small and we see a much less sharp front, with many unoccupied sites behind the leading cell.



(a) $P_m = 0.1, P_p = 5$ (b) $P_m = 5, P_p = 0.1$

Figure 3.3: The results of single discrete simulations, using the algorithm in Section 3.2.1, of a spreading cell front for two different parameter sets show different qualitative behaviour. (a) A high P_p/P_m leads to a sharp front with very few unoccupied sites behind the leading cell. (b) A low P_p/P_m leads to a front with a number of unoccupied sites behind the leading cell. In both cases, we begin with $l \leq 100$ occupied and $l > 100$ unoccupied.

3.2.2 The mean-field approximation

We derive the MFA in a similar way to the previous chapter, but we can no longer assume translational invariance. The average occupancy of the l^{th} lattice site is given by $C(l) \in [0, 1]$. In keeping with the previous chapter, we use k-point distribution functions, $\rho^{(k)}$, to derive the MFA and PWA. The one-point distribution function, $\rho^{(1)}$, gives the averaged occupancy of the site in question. Thus we have

$$\rho^{(1)}(A_l) = C(l), \tag{3.2.2}$$

$$\rho^{(1)}(0_l) = 1 - C(l), \tag{3.2.3}$$

where A_l and 0_l indicate that site l is occupied by A or unoccupied, respectively. We cannot set all $C(l)$ equal to the averaged system density as we did in the previous chapter, due to the average density now varying in space. Instead, we must keep track

of the occupancy probabilities of every site in the system. The 1-point distribution functions will evolve in the following way:

$$\begin{aligned}\frac{\mathrm{d}\rho^{(1)}(A_l)}{\mathrm{d}t} =& \underbrace{\frac{P_m}{2}\left[\rho^{(2)}(A_{l-1},0_l)+\rho^{(2)}(0_l,A_{l+1})-\rho^{(2)}(A_l,0_{l+1})-\rho^{(2)}(0_{l-1},A_l)\right]}_{\text{movement in and out of site } l} \\ &+\underbrace{\frac{P_p}{2}\left[\rho^{(2)}(A_{l-1},0_l)+\rho^{(2)}(0_l,A_{l+1})\right]}_{\text{proliferation into site } l}.\end{aligned} \tag{3.2.4}$$

In the mean-field case, we assume that lattice site occupancies are independent of one another, thus we can rewrite our 2-point distribution functions as the product of the relevant one-point distribution functions:

$$\begin{aligned}\frac{\mathrm{d}\rho^{(1)}(A_l)}{\mathrm{d}t} =& \frac{P_m}{2}\left[\rho^{(1)}(A_{l-1})\rho^{(1)}(0_l)+\rho^{(1)}(0_l)\rho^{(1)}(A_{l+1})\right. \\ &\left.-\rho^{(1)}(A_l)\rho^{(1)}(0_{l+1})-\rho^{(1)}(0_{l-1})\rho^{(1)}(A_l)\right] \\ &+\frac{P_p}{2}\left[\rho^{(1)}(A_{l-1})\rho^{(1)}(0_l)+\rho^{(1)}(0_l)\rho^{(1)}(A_{l+1})\right].\end{aligned} \tag{3.2.5}$$

Rearranging, and expressing the 1-point distribution functions as the averaged site occupancy, we obtain

$$\begin{aligned}\frac{\mathrm{d}C(l)}{\mathrm{d}t} =& \frac{P_m}{2}\left[C(l-1)-2C(l)+C(l+1)\right] \\ &+\frac{P_p}{2}\left[C(l-1)+C(l+1)\right]\left[1-C(l)\right],\end{aligned} \tag{3.2.6}$$

remembering the equations will be modified at the boundaries, where we implement reflecting boundary conditions. In the case with uniform initial conditions, the MFA does not depend on P_m as all P_m-containing terms cancel. However, as we can no longer assume translational invariance, the MFA now depends on P_m as well as P_p. To make predictions using the MFA, we solve the system given by equation (3.2.6) numerically using RK4 [Press et al., 2007] with a constant time step of $\delta t = 0.1$.

Smaller time steps were tested to confirm that this was an appropriate choice. We calculate v and v_f in the same manner as in our discrete model, whereby we track the location of $C(l) = 0.5$. We have initial conditions of the form of equation (3.2.1).

3.2.3 Pair-wise approximation

As we saw in Chapters 1 and 2, it is not always valid to assume that lattice site occupancies are independent. To provide a measure of the occupancy dependence of two given sites we use correlation functions, similar to those defined in equation (2.5.1) in the previous chapter. However, in the previous chapter, the correlation functions did not depend on the exact position of the sites on the lattice, but only on the radial distance between the two sites, ($|\mathbf{l} - \mathbf{m}|$). However, without the assumptions of isotropy and translational invariance, our correlation functions will now also depend on the position of the sites, not just their relative separation. Remember that if lattice site occupancies are independent, we have $F_{\sigma_l,\sigma_m}(l, m) \equiv 1$. As before, the correlation functions can be related by using conservation expressions. For example, we can write the conservation equation

$$\rho^{(2)}(A_l, A_m) + \rho^{(2)}(A_l, 0_m) = \rho^{(1)}(A_l). \tag{3.2.7}$$

Using the definition of a correlation function, we have

$$C(l)C(m)F_{A,A}(l, m) + C(l)[1 - C(m)]F_{A,0}(l, m) = C(l), \tag{3.2.8}$$

and thus we can express one correlation function in terms of another:

$$F_{A,0}(l, m) = \frac{1 - C(m)F_{A,A}(l, m)}{1 - C(m)}. \tag{3.2.9}$$

We rewrite equation (3.2.4) using the correlation functions and conservation equations

to obtain

$$\begin{aligned}\frac{\mathrm{d}C(l)}{\mathrm{d}t} =& \frac{P_m}{2}\left[C(l-1) - 2C(l) + C(l+1)\right] \\ &+ \frac{P_p}{2}\left\{C(l-1)\left[1 - C(l)F(l,l-1)\right] + C(l+1)\left[1 - C(l)F(l,l+1)\right]\right\},\end{aligned} \tag{3.2.10}$$

where we have set $F_{A,A} = F$ for notational simplicity. We note that the movement term does not depend on correlations, as all correlation containing terms cancel [Simpson and Baker, 2011]. However, the proliferative terms do depend on correlations, similarly to what we saw in Chapter 2. For the standard MFA, we assume lattice site occupancies are independent, thus $F(l,m) \equiv 1$ for all l, m, and clearly equation (3.2.10) reduces to equation (3.2.6).

For the PWA, we no longer make the mean-field assumption that the occupancies of pairs of sites are independent, thus we do not set $F(l,m) \equiv 1$ in equation (3.2.10). To determine the evolution of our correlation functions, we turn to the 2-point distribution functions. First we consider where the two sites in question are nearest neighbours:

$$\begin{aligned}\frac{\mathrm{d}\rho^{(2)}(A_l, A_{l+1})}{\mathrm{d}t} =& \underbrace{\frac{P_m}{2}\left[\rho^{(3)}\left(A_{l-1}, 0_l, A_{l+1}\right) + \rho^{(3)}\left(A_l, 0_{l+1}, A_{l+2}\right)\right]}_{\text{movement into sites } l \text{ and } l+1} \\ &- \underbrace{\frac{P_m}{2}\left[\rho^{(3)}\left(0_{l-1}, A_l, A_{l+1}\right) + \rho^{(3)}\left(A_l, A_{l+1}, 0_{l+2}\right)\right]}_{\text{movement out of sites } l \text{ and } l+1} \\ &+ \underbrace{\frac{P_p}{2}\left[\rho^{(3)}\left(A_{l-1}, 0_l, A_{l+1}\right) + \rho^{(3)}\left(A_l, 0_{l+1}, A_{l+2}\right)\right]}_{\text{proliferation from neighbours}} \\ &+ \underbrace{\frac{P_p}{2}\left[\rho^{(2)}\left(0_l, A_{l+1}\right) + \rho^{(2)}(A_l, 0_{l+1})\right]}_{\text{proliferation from each other}}.\end{aligned} \tag{3.2.11}$$

Next, we consider the evolution of the 2-point distribution function where the two

sites in question are not nearest neighbours, thus for $|l-m|>1$:

$$\begin{aligned}\frac{\mathrm{d}\rho^{(2)}(A_l,A_m)}{\mathrm{d}t} =& \underbrace{\frac{P_m}{2}\left[\rho^{(3)}(A_{l-1},0_l,A_m)+\rho^{(3)}(0_l,A_{l+1},A_m)\right]}_{\text{movement into site } l}\\ &\underbrace{-\frac{P_m}{2}\left[\rho^{(3)}(0_{l-1},A_l,A_m)+\rho^{(3)}(A_l,0_{l+1},A_m)\right]}_{\text{movement out of site } l}\\ &\underbrace{+\frac{P_m}{2}\left[\rho^{(3)}(A_l,A_{m-1},0_m)+\rho^{(3)}(A_l,0_m,A_{m+1})\right]}_{\text{movement into site } m}\\ &\underbrace{-\frac{P_m}{2}\left[\rho^{(3)}(A_l,0_{m-1},A_m)+\rho^{(3)}(A_l,A_m,0_{m+1})\right]}_{\text{movement out of site } m}\\ &\underbrace{+\frac{P_p}{2}\left[\rho^{(3)}(A_{l-1},0_l,A_m)+\rho^{(3)}(0_l,A_{l+1},A_m)\right]}_{\text{proliferation into site } l}\\ &\underbrace{+\frac{P_p}{2}\left[\rho^{(3)}(A_l,A_{m-1},0_m)+\rho^{(3)}(A_l,0_m,A_{m+1})\right]}_{\text{proliferation into site } m}.\end{aligned} \tag{3.2.12}$$

In equations (3.2.11) and (3.2.12), we use conservation expressions for the 3-point distribution functions to eliminate some of the 3-point distribution function terms. Where this is not possible, we close using the KSA, equation (2.5.14). We relate the 2-point distribution functions to the correlation functions using equation (2.5.1) to obtain equations for the evolution of the correlation functions. The evolution of $F(l,l+1)$ is given by

$$
\begin{aligned}
F'(l,l+1) &= -F(l,l+1)\left[\frac{C(l+1)'}{C(l+1)}+\frac{C(l)'}{C(l)}\right] \\
&+ \frac{P_m}{2}\left[\frac{C(l-1)}{C(l)}F(l-1,l+1)+\frac{C(l+2)}{C(l+1)}F(l,l+2)-2F(l,l+1)\right] \\
&+ \frac{P_p}{2}\Bigg[\frac{1}{C(l)}+\frac{1}{C(l+1)}-2F(l,l+1) \\
&+ \frac{C(l-1)}{C(l)(1-C(l))}\left[1-C(l)F(l,l+1)\right]\left[1-C(l)F(l,l-1)\right]F(l-1,l+1) \\
&+ \frac{C(l+2)}{C(l+1)(1-C(l+1))}\left[1-C(l+1)F(l,l+1)\right] \\
&\times \left[1-C(l+1)F(l+1,l+2)\right]F(l,l+2)\Bigg], \qquad (3.2.13)
\end{aligned}
$$

where prime denotes differentiation with respect to time. For any general distance, where $|l-m|>1$, we obtain

$$
\begin{aligned}
F'(l,m) &= -F(l,m)\left[\frac{C(l)'}{C(l)}+\frac{C(m)'}{C(m)}\right]+\frac{P_m}{2}\left[\frac{C(l-1)}{C(l)}F(l-1,m)\right. \\
&\left.+\frac{C(l+1)}{C(l)}F(l+1,m)+\frac{C(m-1)}{C(m)}F(l,m-1)+\frac{C(m+1)}{C(m)}F(l,m+1)-4F(l,m)\right] \\
&+ \frac{P_p}{2}\Bigg[\frac{C(l-1)}{C(l)[1-C(l)]}\left[1-C(l)F(l,l-1)\right]\left[1-C(l)F(l,m)\right]F(m,l-1) \\
&+ \frac{C(l+1)}{C(l)[1-C(l)]}\left[1-C(l)F(l,l+1)\right]\left[1-C(l)F(l,m)\right]F(m,l+1) \\
&+ \frac{C(m+1)}{C(m)[1-C(m)]}\left[1-C(m)F(l,m)\right]\left[1-C(m)F(m,m+1)\right]F(l,m+1) \\
&+ \frac{C(m-1)}{C(m)[1-C(m)]}\left[1-C(m)F(l,m)\right]\left[1-C(m)F(m,m-1)\right]F(l,m-1)\Bigg]. \\
& \qquad (3.2.14)
\end{aligned}
$$

We solve these equations numerically using an RK4 algorithm, the FORTRAN code which we modified for our purposes was kindly provided by Professor Matthew Simpson. We use a constant time step, δt, and for each set of parameters, we test smaller time steps to confirm that the results are visually indistinguishable. We calculate v and v_f in the same way as for the MFA, and we again have reflecting boundary conditions and initial conditions of the form of equation (3.2.1). As discussed by

Simpson and Baker [2011], initial conditions such as these lead to difficulties as the correlation functions are unbounded when $C(l) = 1$ or $C(l) = 0$. This issue can be resolved by using a hybrid approach whereby we use the PWA in regions where $\epsilon < C < (1 - \epsilon)$ and the MFA elsewhere. In this case, we set $\epsilon = 1 \times 10^{-10}$. We also need to choose a truncation value for $F(l, m)$ with $m = l + 1, l + 2, l + 3, ..., l + M$. We truncate at $M = 10$, and test higher values of M to confirm that our truncation choice does not significantly affect our results.

3.2.4 One-hole approximation

The OHA estimates the asymptotic front speed in situations where it is valid to assume there is only one hole (unoccupied lattice site) behind the front. In the moment dynamics framework, we consider occupied sites, whereas now we are approaching the problem from the other angle, focussing on unoccupied lattice sites. In this section we show how to calculate v_f, the asymptotic front speed, using the OHA, relaxing the assumption made in previous work that $P_p + P_m = 1$ [Callaghan et al., 2006]. We know that if $P_m = 0$, then $v_f = P_p/2$ and there will be no holes behind the moving front. For P_m small, there will be a small number of holes behind the front and we will need a correction factor to calculate v_f. We define the one-hole states, whereby the hole is in the n^{th} position behind the leading cell, in the following way:

- $|0\rangle$ = (...11111000...);
- $|1\rangle$ = (...11101000...);
- $|2\rangle$ = (...11011000...);
- $|3\rangle$ = (...10111000...), etc.

Note that we always define states in a frame moving with respect to the front. We only allow transitions between these states. The transitions and their associated rates $W_{ij} \equiv W(|i\rangle \rightarrow |j\rangle)$ are

- $W_{00} = \underbrace{P_p/2,}_{\text{proliferation forwards}}$
- $W_{01} = \underbrace{P_m/2,}_{\text{migration forwards}}$
- $W_{10} = \underbrace{P_p}_{\text{proliferation of cells either side of hole}} + \underbrace{P_m/2,}_{\text{migration of end cell backwards}}$
- $W_{n0} = \underbrace{P_p,}_{\text{proliferation of cells either side of hole}}$
- $W_{n,n-1} = \underbrace{P_m/2,}_{\text{migration from ahead of hole}} \quad n > 1,$
- $W_{n,n+1} = \underbrace{P_p/2}_{\text{proliferation of end cell}} + \underbrace{P_m/2,}_{\text{migration from behind hole}} \quad n \geq 1.$

A schematic of some of the possible transitions between states can be seen in Figure 3.4.

Next, we consider the probability, p_n, of being in state n. As we are only considering the steady-state behaviour, we know that transitions in and out of a state must be equal, leading to the following equations:

$$p_0 W_{01} = \sum_{n=1}^{\infty} p_n W_{n0}, \tag{3.2.15}$$

$$p_1(W_{10} + W_{12}) = p_0 W_{01} + p_2 W_{21}, \tag{3.2.16}$$

$$p_n(W_{n0} + W_{n,n-1} + W_{n,n+1}) = p_{n-1} W_{n-1,n} + p_{n+1} W_{n+1,n}, \quad n > 1. \tag{3.2.17}$$

Using the transition rates, we can rewrite these as

$$\frac{P_m}{2} p_0 = \left(P_p + \frac{P_m}{2}\right) p_1 + P_p \sum_{n=2}^{\infty} p_n, \tag{3.2.18}$$

$$\left(\frac{3P_p}{2} + P_m\right) p_1 = \frac{P_m}{2} p_0 + \frac{P_m}{2} p_2, \tag{3.2.19}$$

$$\left(\frac{3P_p}{2} + P_m\right) p_n = \frac{P_p + P_m}{2} p_{n-1} + \frac{P_m}{2} p_{n+1}, \quad n > 1. \tag{3.2.20}$$

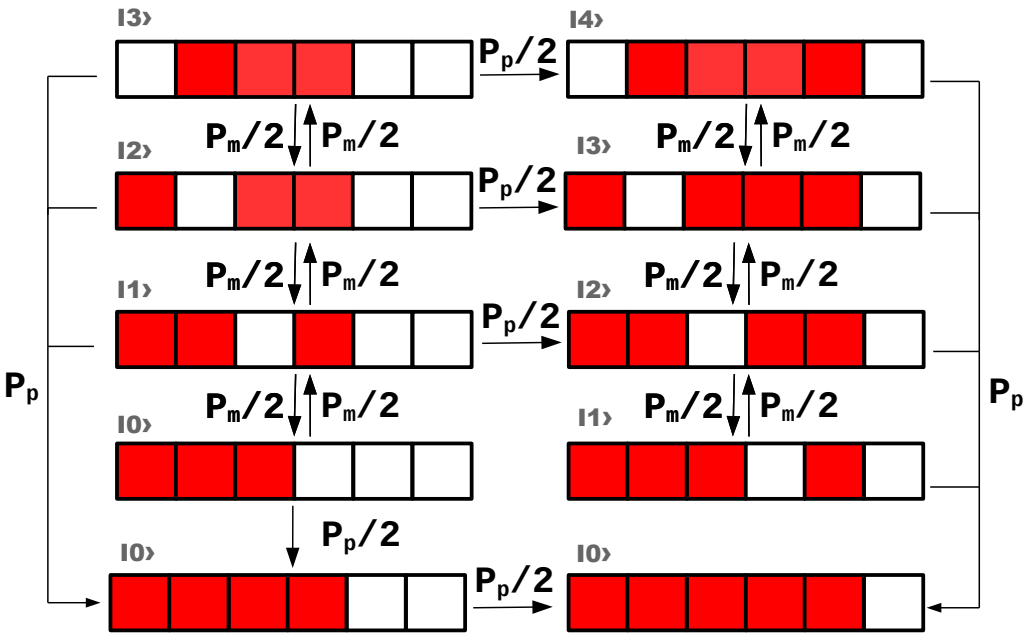


Figure 3.4: Some of the possible transitions between states (upper left corner for each depiction) in the OHA, and their associated rates (along each arrow). Occupied sites are red, whilst empty sites are white. Note that we restrict our transitions to those that only allow one hole (unoccupied site) to be present behind the leading edge.

For $n > 1$, we follow the approach of Callaghan et al. [2006] and assume the ansatz $p_n = a^{n-1}p_1$. We insert this ansatz into equation (3.2.20) to obtain

$$\frac{P_m}{2}a^2 - \left(\frac{3P_p}{2} + P_m\right)a + \frac{P_p + P_m}{2} = 0. \tag{3.2.21}$$

Solving this quadratic in a gives

$$a = \frac{3P_p/2 + P_m \pm \sqrt{9P_p^2/4 + 2P_pP_m}}{P_m}. \tag{3.2.22}$$

We know that as n increases p_n decreases, therefore we take only the negative square root. Assuming that P_m is small relative to P_p, we can use a series expansion to

obtain

$$a = \frac{1}{3} + \frac{4}{27}\left(\frac{P_m}{P_p}\right) - \frac{16}{243}\left(\frac{P_m}{P_p}\right)^2 + \frac{80}{2187}\left(\frac{P_m}{P_p}\right)^3 + \mathcal{O}\left(\frac{P_m}{P_p}\right)^4. \tag{3.2.23}$$

Callaghan et al. [2006] enforced $P_p + P_m = 1$ throughout their derivation. Note that even if we enforce $P_p + P_m = 1$, this is different from the expression derived by Callaghan et al. [2006]. This is because we have relaxed this assumption earlier, thus leading to a slightly different expression for our overall approximation.

We substitute $p_2 = ap_1$ into equation (3.2.19) and use $\sum_{n=2}^{\infty} p_n = 1 - p_0 - p_1$ in equation (3.2.18) resulting in a pair of algebraic equations for p_0 and p_1. We solve these simultaneous equations to obtain

$$p_0 = \frac{3P_pP_m + 4P_p^2 - P_m\sqrt{P_p(9P_p + 8P_m)}}{2(P_pP_m - P_m^2 + 2P_p^2)}, \tag{3.2.24}$$

and

$$p_1 = \frac{6P_p^2 + 7P_pP_m - P_m\sqrt{P_p(9P_p + 8P_m)} + 2P_p\sqrt{P_p(9P_p + 8P_m)}}{2(P_pP_m - P_m^2 + 2P_p^2)}. \tag{3.2.25}$$

Knowing p_0, it is possible for us to approximate v_f for small P_m using a correction factor. We know that if holes are opening up behind the front, v_f will be smaller than if there are no holes behind the front. When $P_m = 0$, $v_f = P_p/2$. For non-zero P_m, there will sometimes be movement forward at a rate $P_m/2$, thus, if there are no holes, $v_f = (P_p + P_m)/2$. There will also be movement back into holes thus we reduce the front speed by taking into account those agents moving back into holes:

$$v_f = \frac{P_p + P_m}{2} - \frac{P_m p_1}{2}. \tag{3.2.26}$$

Performing a series expansion on p_1, assuming P_m is relatively small, we thus obtain the following approximate expression for the front speed:

$$v_f = \frac{P_p + P_m}{2} - \frac{1}{6}\frac{P_m^2}{P_p} + \frac{5}{54}\frac{P_m^3}{P_p^2} - \frac{340 - 162P_p}{3888}\frac{P_m^4}{P_p^3} + \mathcal{O}(P_m^5). \tag{3.2.27}$$

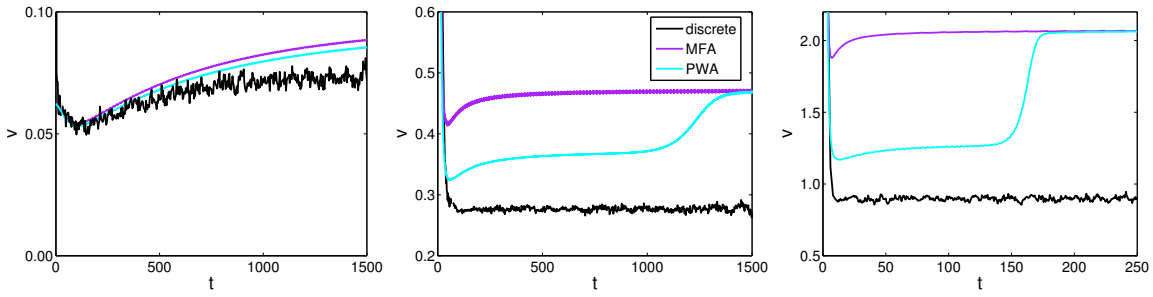
Therefore, we estimate v_f to be 0.498 when $P_p = 0.9$ and $P_m = 0.1$, agreeing with the results of Callaghan et al. [2006].

3.3 Results

In this section we examine the behaviour of our three derived methods (MFA, PWA and OHA), comparing their predictions to the averaged discrete behaviour across a range of parameters and timescales. We begin with the transient behaviour before considering the asymptotic results.

3.3.1 Transient behaviour

In practice, we must wait a sufficient duration of time to observe the asymptotic speed, which in theory is only fully established as $t \to \infty$. In Figure 3.5, we see this for the cases in which we can examine the transient behaviour: the discrete, MFA and PWA. The OHA is only capable of predicting the asymptotic front speed, which can be problematic because, as mentioned earlier, the asymptotic speed may not always be reached in reality, depending on the experimental conditions [Liang et al., 2007; Rodriguez et al., 2005; Valster et al., 2005; Young et al., 2012]. We see, in Figure 3.5, that the time taken to reach the asymptotic travelling front speed varies depending on the model chosen, and the parameter values. For sufficiently low P_p/P_m, the MFA and PWA are both in suitable agreement with the averaged discrete results. For larger P_p/P_m, the discrete model reaches the asymptotic speed noticeably faster than the



(a) $P_p/P_m = 5 \times 10^{-3}$ (b) $P_p/P_m = 10^{-1}$ (c) $P_p/P_m = 1$

Figure 3.5: The transient front speed, v, varies over time, eventually reaching its asymptotic travelling speed, v_f. We see this behaviour in the three different cases for which we can examine the transient behaviour: the discrete (using the algorithm in Section 3.2.1), MFA (equation (3.2.6)) and PWA (using the equations in Section 3.2.3). The behaviour depends on the parameters chosen, and we compare three P_p/P_m ratios (for all cases $P_m = 1$ and P_p is varying). (a) For low P_p/P_m, we notice that all three methods are in relatively good agreement. (b), (c) For larger P_p/P_m, the MFA and PWA begin to deviate from the averaged discrete results, generally tending to a higher asymptotic speed. In all cases, the MFA and PWA eventually reach the same travelling speed, although the PWA takes significantly longer. The time taken for the PWA to reach the same speed as the MFA decreases as P_p/P_m increases, as we see by comparing (b) and (c). In this case, $x = 50$ in equation (3.2.1) for the initial conditions.

MFA. The PWA takes significantly longer and changes very slowly for a long period before more rapidly adjusting to the asymptotic speed. The asymptotic speed, v_f, is identical for the MFA and PWA in all cases, with this value being higher than the averaged discrete result for large P_p/P_m. The asymptotic speed is identical for the MFA and PWA due to the fact that we have used a hybrid PWA whereby we use the MFA in regions where $C < \epsilon$ and $C > 1 - \epsilon$. Changing ϵ shifts the point at which the asymptotic speed is reached; a larger value of ϵ leads to the asymptotic speed being reached more quickly (Figure 3.6).

Additionally, we examine the effects of altering the steepness of the initial conditions by varying x in equation (3.2.1). Altering the initial conditions, as long as compact support is maintained, does not affect the asymptotic results. However, different initial conditions have an impact on the transient behaviour. We see, in

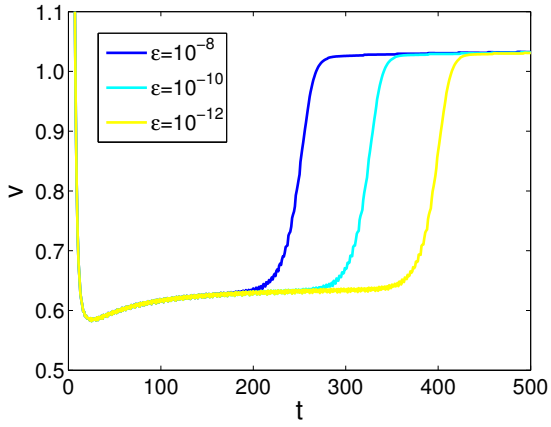
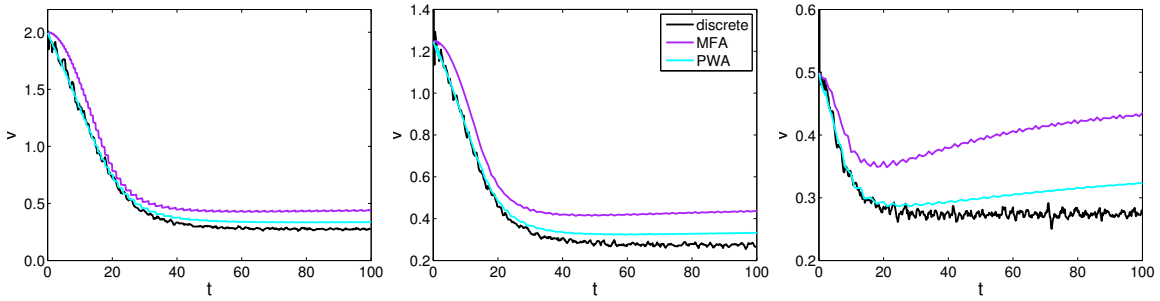


Figure 3.6: Changing ϵ in the PWA (given by the equations in Section 3.2.3) changes the point at which the asymptotic speed is reached. A higher value of ϵ leads to the speed being reached more rapidly. The initial and final behaviour for all values of ϵ are visually indistinguishable. The parameters in this case are $P_m = P_p = 0.5$.

Figure 3.7, that a steeper ramp (which corresponds to a larger value of x in equation (3.2.1)) leads to a lower initial speed. The MFA and PWA also lie slightly closer to the averaged discrete behaviour at early times when the initial conditions are steeper.

We compare the transient density profiles at various times in Figure 3.8. From this we see that the PWA better predicts the averaged discrete behaviour than does the MFA. Thus, whilst the asymptotic behaviour might be the same, the PWA is more accurate at predicting the average transient behaviour of the system when P_p is sufficiently low relative to P_m. We also note, by comparing the top and bottom rows of Figure 3.8, that a steeper initial slope leads to the MFA and PWA better approximating the averaged discrete behaviour in the transient region, agreeing with what we observed in Figure 3.7.



(a) $x = 20$ (b) $x = 50$ (c) $x = 80$

Figure 3.7: Adjusting the steepness of the initial conditions affects the transient behaviour. We see that a steeper slope (given by a higher value of x in our initial conditions, equation (3.2.1)) leads to the initial speed being lower, and the PWA (equations in Section 3.2.3) and MFA (equation (3.2.6)) providing a slightly closer agreement with the averaged discrete behaviour. In all cases, we see that the PWA provides a significant improvement to the MFA at predicting the averaged discrete behaviour in the transient region. The parameters in this case are $P_m = 1, P_p = 0.1$.

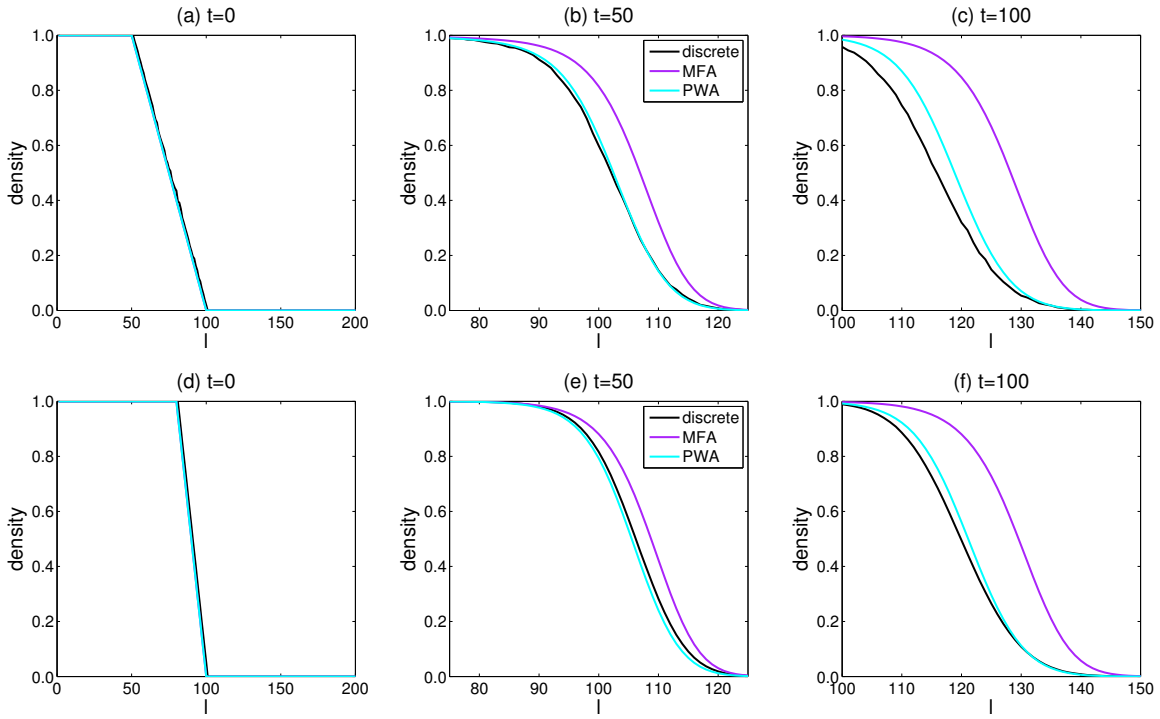


Figure 3.8: Density plots for early times show that the PWA (using the equations in Section 3.2.3) provides a closer approximation of the averaged discrete behaviour than does the MFA, equation (3.2.6). The parameters in this case are $P_m = 1, P_p = 0.1$. (a)-(c) Time evolution for initial conditions whereby $x = 50$. (d)-(f) Time evolution for initial conditions whereby $x = 80$. We see that a steeper initial ramp leads to improved performance of the MFA and PWA at approximating the transient averaged discrete behaviour.

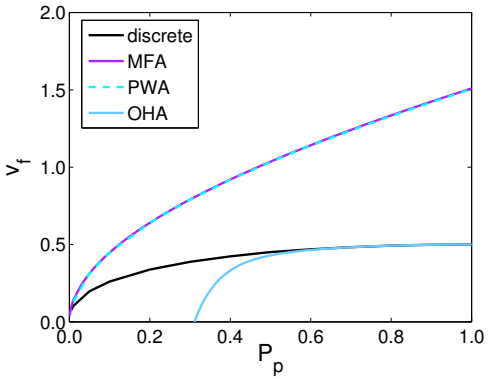
3.3.2 Asymptotic behaviour

As $t \to \infty$, all cases predict a travelling front with a constant speed. We first look at the parameter regime where $P_p + P_m = 1$, as done by Callaghan et al. [2006], in Figure 3.9(a). We see that the OHA provides a good estimate of v_f for high enough P_p, but diverges when P_p is reduced beyond a certain level. We also look at the average number of holes behind the front in the discrete case (Figure 3.9(b)) for the same parameter range, noting that this corresponds well with the predictive power of the OHA: as we move beyond one hole on average, the OHA breaks down. The break down of the OHA can also be attributed to the assumption that P_m is small relative to P_p which ceases to hold as we increase P_m and decrease P_p. The MFA prediction for v_f (which is identical to the PWA prediction) agrees well with the averaged discrete case for low P_p but is not a good predictor when P_p is large. For a region in the middle, neither the MFA nor the OHA provides a suitable approximation to the averaged discrete behaviour.

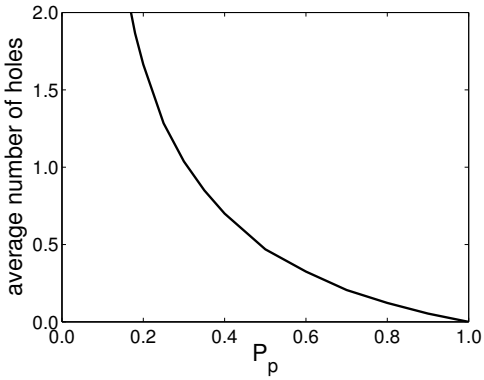
Additionally, we relax the assumption made by Callaghan et al. [2006], no longer requiring that $P_p + P_m = 1$. This allows us to look at a more relative measure, the ratio of proliferation to migration rates, P_p/P_m. We keep $P_m = 1$ and allow P_p to vary (Figure 3.10(a)), observing similar results to before, whereby the OHA performs well when P_p is sufficiently large relative to P_m, whilst the MFA performs well when P_p is sufficiently small relative to P_m. We again look at the number of holes behind the front in the averaged discrete case in Figure 3.10(b), noting that the OHA fails as this is increased beyond unity.

3.4 Discussion

Mathematical models are often used in conjunction with experimental data to examine moving cell fronts, in the hope of determining information such as the mecha-

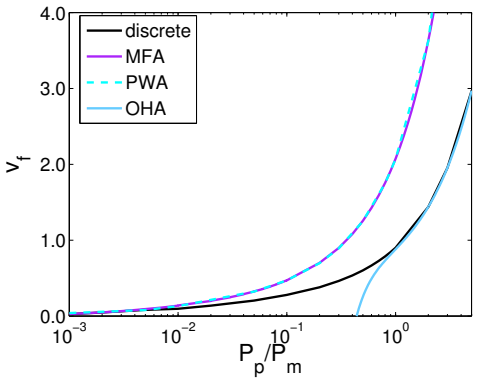


(a) Asymptotic speed

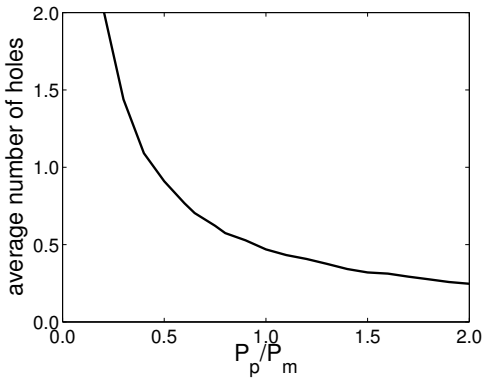


(b) Holes behind the front

Figure 3.9: (a) The asymptotic travelling wave speed varies with the parameters. In this case, we have set $P_m + P_p = 1$. As the proliferation rate, P_p, is decreased and the movement rate, P_m, is increased, the MFA (equation (3.2.6)) and PWA (equations in Section 3.2.3), which are identical, agree more closely with the averaged discrete behaviour (using the algorithm in Section 3.2.1). As P_p is increased, the OHA (equation (3.2.27)) becomes a good method for approximating the averaged discrete behaviour. (b) The average number of holes at $t = 1000$ for the discrete case, showing that the OHA fails when we have more than one hole on average.



(a) Asymptotic speed



(b) Holes behind the front

Figure 3.10: (a) The asymptotic travelling wave speed varies with P_p/P_m. We see that the OHA, equation (3.2.27), performs best at approximating the averaged discrete behaviour when P_p/P_m is sufficiently large, whilst the MFA, equation (3.2.6), and PWA (using the equations in Section 3.2.3), which are identical, perform best when P_p/P_m is sufficiently small. We have used a logarithmic scale for P_p/P_m to emphasise the excellent agreement for the MFA and PWA for small P_p/P_m. (b) We again look at the average number of holes behind the front confirming that the OHA does not provide as good an approximation as this increases beyond unity.

	Small P_p/P_m		Intermediate P_p/P_m		Large P_p/P_m	
	Short	Long	Short	Long	Short	Long
MFA	✓	✓	✗	✗	✗	✗
PWA	✓	✓	✗	✗	✗	✗
OHA	✗	✗	✗	✗	✗	✓
Discrete	✓	✓	✓	✓	✓	✓

Table 3.1: A summary of which methods can be used under specific circumstances, where long and short refer to the asymptotic and transient behaviour, respectively. The best method for given conditions is highlighted in green.

nisms driving the movement of the front. We have compared different methods for modelling the transient and asymptotic moving cell front behaviour in a range of parameter space. These results are summarized in Table 3.1. Many models in the past have focussed on predicting the asymptotic behaviour, whereby a front travelling at constant speed has been established. Whilst this is appropriate in some cases, there are many experimental situations in which the transient behaviour is key. Not only does the final behaviour take some time to be realised in practice, some experiments are performed in such a manner that we may never observe the final behaviour. For instance, in a scratch or punch-hole assay, the cells will be encroaching from multiple directions, resulting in the fronts interacting and the unoccupied region closing before the asymptotic behaviour can be reached [Liang et al., 2007; Rodriguez et al., 2005; Valster et al., 2005]. A major drawback of the OHA is its inability to predict the transient behaviour; it is only able to predict the asymptotic speed. The MFA and PWA can both be used for transient data under certain conditions, with the PWA giving improved results thus making it preferable for transient behaviour.

For asymptotic behaviour, the method best suited to a given situation depends on the relative rates of movement and proliferation. In regions where P_p/P_m is small ($P_p/P_m < 0.01$), we have a shallow front [Swanson et al., 2003] with many holes behind the leading cell, thus the OHA diverges substantially from the averaged discrete results and cannot be successfully used. The MFA and PWA are better suited to

modelling cells in this region of parameter space. As these two methods give the same result asymptotically, it is most sensible to use the MFA for asymptotic predictions as this is a far simpler model to implement. In the experimental results in Figure 3.1, we notice a large number of holes behind the front, thus we expect this cell line [Simpson et al., 2007b; Todaro and Green, 1963] to be best approximated by the MFA asymptotically, and by the PWA in the transient region.

For cell lines where P_p/P_m is relatively large ($P_p/P_m > 0.8$), there will be fewer holes behind the sharp front [Swanson et al., 2003], thus the OHA is the method best suited to asymptotic predictions in this region, as indicated by its excellent agreement with the averaged discrete results. The MFA and PWA do not perform as well in this region. We expect cell lines with low rates of migration relative to proliferation to lie in this region of parameter space. For example, the pair-wise model developed for uniform initial conditions described in the previous chapter [Baker and Simpson, 2010] was used to determine P_p and P_m for a breast cancer cell line, MDA MB 231 [Simpson et al., 2013a]. With $P_p = 0.069$ and $P_m = 0.04$, we have a ratio, $P_p/P_m = 1.7$, which lies within the range for which the OHA is best suited for asymptotic predictions. Thus for MDA MB 231 and similar cell lines, we should use the OHA if we need information about the asymptotic behaviour.

For intermediate P_p/P_m ($0.01 < P_p/P_m < 0.8$), none of the methods discussed here provide a good approximation to the averaged discrete results in the long term, and the PWA becomes less successful in the transient region. Thus developing a model that accurately approximates the averaged discrete behaviour in this region requires further investigation. By extending the OHA to higher numbers of holes, as done by Callaghan et al. [2006], we can improve upon this approximation, but at an increasing level of complexity for every additional hole added. Similarly, the PWA can be extended to triplets and so forth to better predict the transient behaviour in a wider range of parameter space. Again, this becomes increasingly complex as more

information is incorporated.

In future, one could consider the inclusion of death in models of advancing cell fronts. It is fairly straightforward to include death in the MFA and PWA, and this has been done in previous work [Simpson and Baker, 2011]. The OHA can also be extended to include death. However, if death rates are too high, relative to proliferation, we expect the OHA to break down due to there being a large number of holes behind the front. Additionally, we have only considered a 1D problem but the method could be extended to higher dimensions in future in order to be more biologically relevant. Some work has been done on predicting the asymptotic front speed by following the leading cell in a similar way to the OHA for 2D ecological systems [Ellner et al., 1998].

3.5 Conclusion

In this chapter, we have demonstrated the relative merits of the MFA, PWA and OHA in comparison with averaged discrete results for transient and asymptotic behaviour in a broad range of parameter space. Our results demonstrate that it is essential to choose the most appropriate modelling strategy for a given biological system, otherwise inaccurate estimations and predictions may result, which could have serious consequences. In the previous chapter, we discussed a correlations method for the case with uniform initial conditions. In this chapter, we looked at the same method but with non-uniform initial conditions, which forces us to relax the assumption of translational invariance. We see that the system becomes far more unwieldy when we are faced with non-uniform initial conditions as we now have a set of equations to solve for every lattice site. Thus if we have 100 lattice sites in 1D, and truncate at a distance of 10, we will have 11 equations for every site which gives 1100 equations overall. As we observed in the previous chapter, higher dimensions lead to non-uniform equations

which are time-consuming to encode. This would become even more problematic if combined with non-uniform initial conditions. Thus, in this chapter, we have only examined the 1D case. It is clear that, if this method is to be more readily used for a broader range of applications, we need to provide some simplifications.

In the next chapter, we develop a simplified method for including spatial correlations in a birth-migration-death process with uniform initial conditions. This allows the model to be more readily applied and extended to more complex scenarios. We show that the method provides good agreement with the correlations model presented in the previous chapter, again showing its improved performance at approximating the averaged discrete behaviour in comparison to the MFA. In Chapter 2 we did not examine low levels of movement relative to proliferation. We have noticed in this chapter that this is an area of parameter space in which the correlations model begins to break down. It is worth investigating if low motility relative to proliferation causes a similar break down in the case with uniform initial conditions. Therefore, in the next chapter, we examine this region of parameter space using the simplified model which we will first develop.

Chapter 4

Simplifying the one-species correlations model

In the previous chapters, we have demonstrated the importance of using methods that go beyond the mean-field when modelling biological phenomena. In Chapter 2, we discussed an existing method for including spatial correlations in mean-field models with spatially uniform initial conditions. We demonstrated that incorporating additional spatial information led to successful approximations of the individual behaviour, without the computational cost of simulating every agent. However, we found that the model was not straightforward to implement in 2D and 3D, due to the irregular distances in the system of ODEs used to represent the correlation functions. In Chapter 3, whereby we dealt with non-uniform initial conditions, we only demonstrated the method in 1D because including correlations in higher dimensions for non-uniform initial conditions is prohibitively complex. The systems we have discussed so far have been fairly simple, involving only a single-species birth-migration-death process, and most biological systems will be significantly more complex. For correlations models to be more readily used for modelling a wide range of biological systems, it is essential that they are not excessively difficult to implement. Thus it would be useful to

simplify correlation models as much as feasibly possible. In this chapter, we provide a simplification of the method discussed in Chapter 2. We do so by approximating the system of correlation ODEs as a PDE, which we find to provide an excellent approximation to the correlations ODE model, thus also successfully predicting the averaged discrete results. Additionally, we investigate a wider range of parameter space than was done previously, considering cases where migration rates are low relative to those of proliferation.

4.1 Introduction

Whilst including spatial correlations in our population models leads to better agreement with the averaged discrete results, these models can be challenging to implement. We saw, in Chapter 2, that including correlation functions in 2D and 3D leads to irregular systems of ODEs due to the radial distances between lattice sites not following an easily prescribed pattern. This means that these equations are difficult to implement in practice, making it challenging to extend this method beyond very simple cases. In an effort to tackle this problem, studies have been carried out to provide a PDE description of nearest-neighbour two-point correlations for an invasion problem with adhesion and contact interactions [Ascolani et al., 2013]. However, spatial correlations are not necessarily confined to nearest-neighbour pairs, for which this approach cannot account. A PDE description of spatial correlations exists for off-lattice point-particle (non-excluding) models [Young et al., 2001], however its parameters must be obtained from averaged discrete simulations, thus it cannot stand alone. As such, inclusion of the effects of spatial correlations into such models in a tractable way remains an open challenge.

In this chapter, we derive a simplified correlations model for a one-species system containing agents undergoing proliferation, movement and death on a regular d-

dimensional lattice with volume exclusion and uniform initial conditions. This model is based on that in Chapter 2 [Baker and Simpson, 2010], however we derive a PDE to represent the correlations, rather than using a system of ODEs. This leads to a more tractable method for including spatial correlations in systems such as these. We confirm that our representation agrees with the full ODE system and compare results from this model to those obtained from the MFA and averaged discrete simulations over a range of parameter space.

Additionally, we investigate a wider parameter regime than was examined previously for uniform initial conditions, taking into account regions of parameter space where motility is reduced relative to proliferation. This phenomenon has been observed in some forms of cancer where cells are proliferating rapidly but do not have high levels of motility [Smolle et al., 1992]. Further, this region of parameter space may also be important for modelling plant dynamics, where movement generally does not feature although proliferation tends to be more widespread due to the effects of dispersal [Czaran and Bartha, 1992; Silvertown et al., 1992]. We find that our simplified model provides significantly improved agreement with the averaged discrete behaviour, compared to the performance of the MFA, in a wide range of parameter space. However, when migration and death rates are high relative to those of proliferation, we see that the correlations model is less successful at predicting the averaged discrete behaviour. However it does still provide an improvement on the MFA.

We begin by deriving the correlations PDE in 1D (Section 4.2), 2D (Section 4.3) and 3D (Section 4.4), before testing the model in Section 4.5. Additionally, we discuss some insights gained by this formulation in Section 4.6, and we then investigate our model across a range of parameter space in Section 4.7 before concluding the chapter.

4.2 Derivation in one dimension

The model described in Chapter 2 provides much better agreement with the averaged discrete behaviour than does the MFA. However, as we have demonstrated, it is not ideal to implement. In this section, we derive a PDE formulation for our system of correlation ODEs in 1D. As a reminder, the correlations ODEs we derived in Chapter 2 have the following form in 1D for $x > \Delta$:

$$\begin{aligned}\frac{\mathrm{d}F(x)}{\mathrm{d}t} = &P_m[F(x-\Delta) - 2F(x) + F(x+\Delta)] - 2P_pF(x)[1 - c_AF(\Delta)] \\ &+ \frac{P_p}{1-c_A}[1 - c_AF(\Delta)][1 - c_AF(x)][F(x-\Delta) + F(x+\Delta)].\end{aligned} \tag{4.2.1}$$

Remember, from Chapter 2, that death terms do not appear in our correlation equations, thus this derivation is valid for systems with and without death. We obtain a PDE for our correlation equations by performing a Taylor expansion about x in equation (4.2.1), where Δ is a small parameter representing the lattice size. Biologically, Δ corresponds to the diameter of a cell, thus we anticipate our approximation to be valid on length scales greater than that of a single lattice site. Additionally, the expression for $|\mathbf{l} - \mathbf{m}| = \Delta$ has a different form from those for $|\mathbf{l} - \mathbf{m}| > \Delta$ because volume exclusion prevents proliferation or movement into site l. Thus we explicitly retain our expression for $|\mathbf{l} - \mathbf{m}| = \Delta$, with this acting as a boundary condition on the PDE we shall now derive. Expanding equation (4.2.1), keeping terms up to second order, gives us

$$\begin{aligned}\frac{\partial F(x)}{\partial t} =& P_m\left[F(x) - \Delta\frac{\partial F(x)}{\partial x} + \frac{\Delta^2}{2}\frac{\partial^2 F(x)}{\partial x^2} - 2F(x) + F(x) + \Delta\frac{\partial F(x)}{\partial x} + \frac{\Delta^2}{2}\frac{\partial^2 F(x)}{\partial x^2}\right] \\ & - 2P_p F(x)[1 - c_A F(\Delta)] + \frac{P_p}{1 - c_A}[1 - c_A F(\Delta)][1 - c_A F(x)] \\ & \times\left[F(x) - \Delta\frac{\partial F(x)}{\partial x} + \frac{\Delta^2}{2}\frac{\partial^2 F(x)}{\partial x^2} + F(x) + \Delta\frac{\partial F(x)}{\partial x} + \frac{\Delta^2}{2}\frac{\partial^2 F(x)}{\partial x^2}\right] \\ & + \mathcal{O}(\Delta^3).\end{aligned} \tag{4.2.2}$$

Simplifying this expression, we obtain

$$\begin{aligned}\frac{\partial F(x)}{\partial t} =& P_m\Delta^2\frac{\partial^2 F(x)}{\partial x^2} - 2P_p F(x)[1 - c_A F(\Delta)] \\ & + \frac{P_p}{1 - c_A}\left[1 - c_A F(\Delta)\right]\left[1 - c_A F(x)\right]\left[2F(x) + \frac{\Delta^2}{2}\frac{\partial^2 F(x)}{\partial x^2}\right], \qquad x > \Delta.\end{aligned} \tag{4.2.3}$$

Generally, in obtaining a PDE, we now take the limit as $\Delta \to 0$ [Codling et al., 2008; Othmer et al., 1988; Simpson et al., 2009a, 2010a; Stevens and Othmer, 1997]. At this point, it is important to consider how P_m will change if Δ is modified. P_m has been chosen as the rate at which an agent will attempt to move to an unoccupied neighbouring lattice site. We notice that if P_m does not scale with Δ^2, we will lose our movement terms entirely thus we expect $P_m\Delta^2$ to remain constant, in keeping with other PDE derivations that do not incorporate correlations [Codling et al., 2008; Othmer et al., 1988; Simpson et al., 2009a, 2010a; Stevens and Othmer, 1997]. This can be expressed as

$$\hat{P}_m = P_m\Delta^2. \tag{4.2.4}$$

In our discrete simulations we have taken $\Delta = 1$ thus $\hat{P}_m = P_m$. For different values of Δ, $\hat{P}_m$ can be obtained using equation (4.2.4). Upon taking the limit as $\Delta \to 0$, we obtain the PDE

$$\frac{\partial F(x)}{\partial t} = \hat{P}_m \frac{\partial^2 F(x)}{\partial x^2} + 2P_p c_A F(x)\left[1 - c_A F(\Delta)\right]\left[\frac{1 - F(x)}{1 - c_A}\right], \qquad x > \Delta, \quad (4.2.5)$$

with boundary conditions

$$\begin{aligned}\frac{\mathrm{d}F(\Delta)}{\mathrm{d}t} &= P_m[F(2\Delta) - F(\Delta)] - 2P_p F(\Delta)[1 - c_A F(\Delta)] \\ &\quad + \frac{P_p}{c_A}[1 - c_A F(\Delta)] + \frac{P_p}{1 - c_A} F(2\Delta)[1 - c_A F(\Delta)]^2,\end{aligned} \quad (4.2.6)$$

and

$$F(x = x_{max}) = 1. \quad (4.2.7)$$

We explicitly retain our equation for $F(\Delta)$, using equation (4.2.6) as the boundary condition at $x = \Delta$, as our approximation is only valid over length scales greater than a single lattice site. The far-field boundary condition is unity as we expect correlations to decay in the large distance limit. Similarly to the 1D ODE case, we choose a maximum value of x, x_{max}, beyond which we set the correlations to unity. Our discrete simulations will obviously be on a finite lattice, but we assume the lattice is large enough such that $F(x_{max}) = 1$ is a valid approximation. We test that our choice of x_{max} is appropriate by increasing x_{max} in our system of equations and confirming the results are visually identical. We now turn to the derivation in 2D which is not as straightforward as in 1D due to the non-uniform distances between lattice sites.

4.3 Derivation in two dimensions

We have seen that it is fairly straightforward to derive a PDE for the correlation functions in 1D. However, we expect the derivation to be more challenging in 2D as we must deal with non-uniform distances on the lattice. Due to the homogeneous initial conditions our system is translationally invariant, and we assume it is also

radially symmetric, thus we need only consider correlation functions for sites a radial distance s apart, rather than each pair of points on the lattice. However, to allow us to develop a PDE approximation, it is more convenient to express our correlation functions in terms of Cartesian coordinates (x, y in 2D and x, y, z in 3D) for our lattice site at radial distance $|\mathbf{l}-\mathbf{m}|$. This allows us to write all the equations for evolution of the correlation functions, apart from when the distance between sites is one cell, in the same form. We can then use Taylor series expansions to approximate the system of ODEs by a PDE, as we did in 1D. The Cartesian coordinate representation for the correlation functions in 2D, where $x^2+y^2>\Delta$, is given by

$$\begin{aligned}\frac{\mathrm{d}F(x,y)}{\mathrm{d}t} =& \frac{P_m}{2}\left[F(x+\Delta,y)+F(x-\Delta,y)+F(x,y+\Delta)+F(x,y-\Delta)-4F(x,y)\right]\\ &-2P_pF(x,y)[1-c_AF(\Delta)]+\frac{P_p}{2(1-c_A)}\left[1-c_AF(\Delta)\right]\left[1-c_AF(x,y)\right]\\ &\times\left[F(x+\Delta,y)+F(x-\Delta,y)+F(x,y+\Delta)+F(x,y-\Delta)\right].\end{aligned} \tag{4.3.1}$$

We perform a 2D Taylor series expansion of equation (4.3.1) around (x, y) keeping terms up to second order in Δ to obtain

$$\begin{aligned}\frac{\partial F(x,y)}{\partial t} =& \frac{P_m\Delta^2}{2}\left[\frac{\partial^2F(x,y)}{\partial x^2}+\frac{\partial^2F(x,y)}{\partial y^2}\right]-2P_pF(x,y)[1-c_AF(\Delta)]\\ &+\frac{P_p}{2(1-c_A)}\left[1-c_AF(\Delta)\right]\left[1-c_AF(x,y)\right]\\ &\times\left[4F(x,y)+\Delta^2\frac{\partial^2F(x,y)}{\partial x^2}+\Delta^2\frac{\partial^2F(x,y)}{\partial y^2}\right]+\mathcal{O}(\Delta^3).\end{aligned} \tag{4.3.2}$$

We can now return to radial coordinates:

$$\begin{aligned}\frac{\partial F(s)}{\partial t} =& \frac{P_m\Delta^2}{2}\left[\frac{\partial^2F(s)}{\partial s^2}+\frac{1}{s}\frac{\partial F(s)}{\partial s}\right]-2P_pF(s)[1-c_AF(\Delta)]\\ &+\frac{P_p}{2(1-c_A)}\left[1-c_AF(\Delta)\right]\left[1-c_AF(s)\right]\\ &\times\left[4F(s)+\Delta^2\frac{\partial^2F}{\partial s^2}+\Delta^2\frac{1}{s}\frac{\partial F}{\partial s}\right], \qquad s>\Delta,\end{aligned} \tag{4.3.3}$$

where s is the radial distance, $|\mathbf{l}-\mathbf{m}|$, between lattice sites l and m. Upon taking the limit as $\Delta \to 0$, with the same argument governing P_m scaling as before, we obtain the following PDE:

$$\frac{\partial F(s)}{\partial t} = \frac{\hat{P}_m}{2}\left[\frac{\partial^2 F(s)}{\partial s^2} + \frac{1}{s}\frac{\partial F(s)}{\partial s}\right] + 2P_p c_A F(s)\left[1 - c_A F(\Delta)\right]\left[\frac{1-F(s)}{1-c_A}\right], \qquad s > \Delta, \tag{4.3.4}$$

with boundary conditions

$$\begin{aligned}\frac{\mathrm{d}F(\Delta)}{\mathrm{d}t} = {} & \frac{P_m}{2}\left[2F(\sqrt{2}\Delta) + F(2\Delta) - 3F(\Delta)\right] - 2P_p F(\Delta)[1 - F(\Delta)c_A] \\ & + \frac{P_p}{2c_A}\left[1 - c_A F(\Delta)\right] + \frac{P_p}{2(1-c_A)}\left[1 - c_A F(\Delta)\right]^2\left[2F(\sqrt{2}\Delta) + F(2\Delta)\right],\end{aligned} \tag{4.3.5}$$

and

$$F(s = s_{max}) = 1. \tag{4.3.6}$$

Again, we explicitly retain the boundary equation (4.3.5) in our system, since we expect our approximation, equation (4.3.4), to be valid only over length scales greater than a single lattice site. As we discussed for the 1D case, we expect the correlations to decay in the limit of large lattice spacing, thus the correlation functions tend to unity for the far-field boundary condition, and we choose s_{max} accordingly.

We notice that the PDEs in 1D, equation (4.2.5), and 2D, equation (4.3.4), have the same form, with the proliferative term being identical in both cases. The diffusion coefficient is P_m in 1D and $P_m/2$ in 2D, and thus appears to scale with the number of neighbours for each agent.

4.4 Derivation in three dimensions

We derive the PDE in 3D using the same method. For $x^2+y^2+z^2 > \Delta$, we can again write our correlations equations in cartesian form:

$$\begin{aligned}\frac{\mathrm{d}F(x,y,z)}{\mathrm{d}t} =& \frac{P_m}{3}\left[F(x+\Delta,y,z)+F(x-\Delta,y,z)+F(x,y+\Delta,z)\right.\\ &\left.+F(x,y-\Delta,z)+F(x,y,z+\Delta)+F(x,y,z-\Delta)-6F(x,y,z)\right]\\ &-2P_pF(x,y,z)[1-c_AF(\Delta)]+\frac{P_p}{3(1-c_A)}\left[1-c_AF(\Delta)\right]\left[1-c_AF(x,y,z)\right]\\ &\times\left[F(x+\Delta,y,z)+F(x-\Delta,y,z)+F(x,y+\Delta,z)\right.\\ &\left.+F(x,y-\Delta,z)+F(x,y,z+\Delta)+F(x,y,z-\Delta)\right].\end{aligned} \tag{4.4.1}$$

As before, we Taylor expand around (x,y,z) keeping terms up to second order:

$$\begin{aligned}\frac{\partial F}{\partial t} =& \frac{P_m\Delta^2}{3}\left[\frac{\partial^2F}{\partial x^2}+\frac{\partial^2F}{\partial y^2}+\frac{\partial^2F}{\partial z^2}\right]-2P_pF[1-c_AF(\Delta)]\\ &+\frac{P_p}{3(1-c_A)}\left[1-c_AF(\Delta)\right]\left[1-c_AF\right]\\ &\times\left[6F+\Delta^2\frac{\partial^2F}{\partial x^2}+\Delta^2\frac{\partial^2F}{\partial y^2}+\Delta^2\frac{\partial^2F}{\partial z^2}\right]+\mathcal{O}(\Delta^3),\end{aligned} \tag{4.4.2}$$

where $F=F(x,y,z)$ for clarity. As with 1D and 2D, we allow $\Delta\to 0$, remembering that P_m scales with Δ^2, obtaining

$$\begin{aligned}\frac{\partial F}{\partial t} =& \frac{\hat{P_m}}{3}\left[\frac{\partial^2F}{\partial x^2}+\frac{\partial^2F}{\partial y^2}+\frac{\partial^2F}{\partial z^2}\right]-2P_pF[1-c_AF(\Delta)]\\ &+\frac{P_p}{1-c_A}\left[1-c_AF(\Delta)\right]\left[1-c_AF\right]2F.\end{aligned} \tag{4.4.3}$$

Finally, we switch to radial coordinates, noting that our system will be radially symmetric due to the homogeneous initial conditions, to arrive at

$$\frac{\partial F(s)}{\partial t} = \frac{\hat{P}_m}{3}\left(\frac{\partial^2 F(s)}{\partial s^2} + \frac{2}{s}\frac{\partial F(s)}{\partial s}\right) + 2P_p c_A F(s)\left[1 - c_A F(\Delta)\right]\left[\frac{1 - F(s)}{1 - c_A}\right], \qquad s > \Delta, \tag{4.4.4}$$

with boundary conditions

$$\begin{aligned}\frac{\mathrm{d}F(\Delta)}{\mathrm{d}t} =& \frac{P_m}{3}\left[4F(\sqrt{2}\Delta) + F(2\Delta) - 5F(\Delta)\right] - 2P_p F(\Delta)[1 - F(\Delta)c_A] \\ &+ \frac{P_p}{3c_A}\left[1 - c_A F(\Delta)\right] + \frac{P_p}{3(1 - c_A)}\left[1 - c_A F(\Delta)\right]^2\left[4F(\sqrt{2}\Delta) + F(2\Delta)\right],\end{aligned} \tag{4.4.5}$$

and

$$F(s = s_{max}) = 1. \tag{4.4.6}$$

As expected, the 3D PDE has a similar form to those obtained for the 1D and 2D cases, with the proliferative term being identical and the diffusion coefficient now being $P_m/3$, in keeping with the scaling previously observed.

4.5 Testing the PDE approximation

We test our PDE approximation by comparing it to results from solving the system of correlations ODEs. To solve our PDE system numerically, we now have to consider two ODEs (one for the cell density, equation (2.5.5), and one for correlations at the nearest-neighbour distance) and a PDE. In order to solve our PDE numerically we encode a backwards Euler method, and solve the resulting non-linear algebraic equations using the tridiagonal matrix algorithm [Press et al., 2007] with Picard iteration [Zheng and Bennett, 2002]. We use a time step of $\delta t = 0.01$ and a space step of $\delta x = 0.1$. Again, we test smaller values to ensure these are appropriately selected. We terminate our PDE domain at a distance of 20Δ, assuming all sites further apart are independent. We confirm this assumption for all cases by solving

our PDE on larger domains and ensuring visually similar results are obtained. We find that the PDE consistently provides a reliable approximation to the correlations ODEs, as illustrated in Figure 4.1. We note that the agreement improves as the number of dimensions increases. An assumption we have made is that $\Delta \to 0$, which greatly simplifies the PDE but is not physically feasible. Without this assumption, the PDE in 1D would reduce exactly to the system of ODEs if we set our discretisation parameter to our lattice size. Thus, the 1D PDE would agree exactly with the ODE. However, with the assumption that $\Delta \to 0$, we notice that this is not the case (Figure 4.1). However, the agreement in 2D and 3D is excellent and these are the cases for which we require a PDE: the 1D case can easily be coded with the system of ODEs as the distances follow a regular pattern. Thus, we retain our assumption that $\Delta \to 0$, as we are concentrating mainly on the 2D and 3D cases and wish to keep the model as tractable as possible.

4.6 Insights from the PDE

We observe that equations (4.2.5), (4.3.4) and (4.4.4) have the form of reaction-diffusion equations. In this section we focus on equation (4.3.4), the 2D case, but similar insights apply in 1D and 3D. The diffusion and reaction terms are indicated in the following equation:

$$\frac{\partial F(s)}{\partial t} = \underbrace{\frac{\hat{P_m}}{2}\left[\frac{\partial^2 F(s)}{\partial s^2} + \frac{1}{s}\frac{\partial F(s)}{\partial s}\right]}_{\text{diffusion term}} + \underbrace{2P_p c_A F(s)\left[1 - c_A F(\Delta)\right]\left[\frac{1 - F(s)}{1 - c_A}\right]}_{\text{reaction term}}, \qquad s > \Delta. \tag{4.6.1}$$

The diffusion coefficient depends only on P_m and we see that correlations will decay faster by the diffusive term if P_m is increased. This agrees with our intuition, as we would expect higher rates of movement to break up clusters of agents. We observe in Figure 4.2 that increasing P_m for a constant P_p (in the absence of agent death) causes

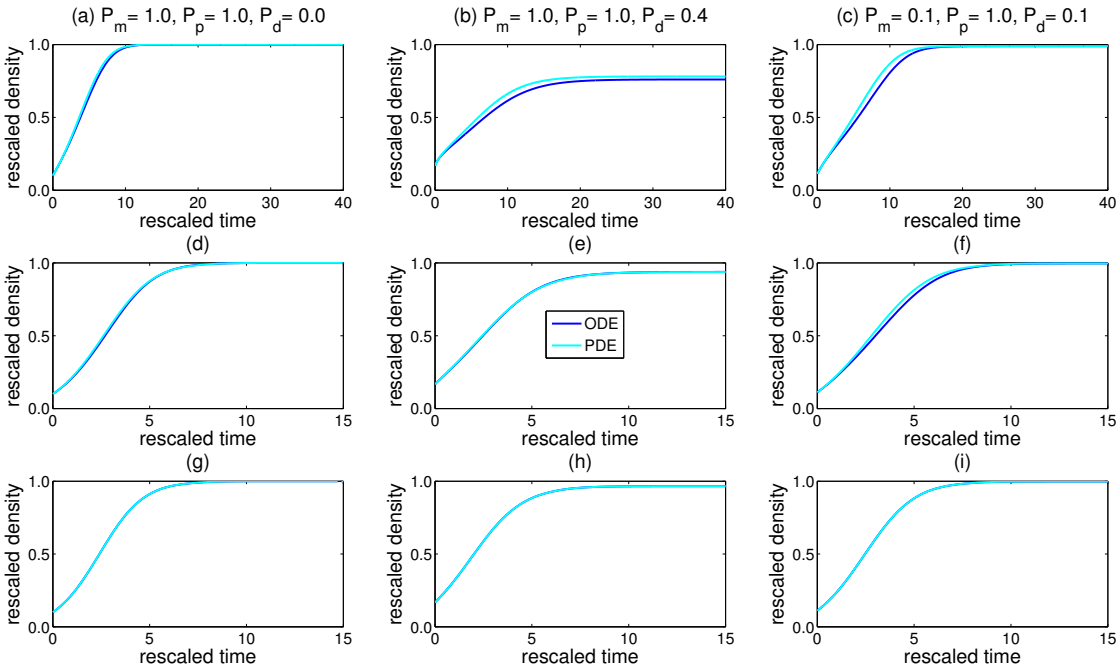


Figure 4.1: Comparing the PDE approximations to the system of ODEs in 1D (top row), 2D (middle row) and 3D (bottom row). We observe that the PDE provides excellent agreement with the system of ODEs in 2D and 3D, being virtually indistinguishable visually. The agreement between the ODE and PDE in 1D is good, but less so than in higher dimensions. As the system of ODEs in 1D is straightforward to solve, this does not pose a significant problem.

correlations to decay faster. When P_m is smaller, the correlation functions have a greater maximum which is to be expected due to the slower dispersal.

Our 'reaction' term depends on the rate of proliferation, however it is unclear at first whether this term is positive or negative. We know that $2P_p c_A$ will always be positive by definition. We next examine the fraction $(1 - F(s))/(1 - c_A)$, noting that the denominator is always positive as $c_A \leq 1$ by definition. The correlation function, $F(s)$, is unity if sites are uncorrelated. If sites are positively correlated $F(s) > 1$, and if sites are negatively correlated $F(s) < 1$. In our system, sites cannot become negatively correlated as this would only happen if, for instance, cells were more likely to die when they were surrounded by other cells. Thus $F(s) \geq 1$. Therefore, we expect $(1 - F(s))/(1 - c_A)$ to be non-positive. Finally, the remaining component of

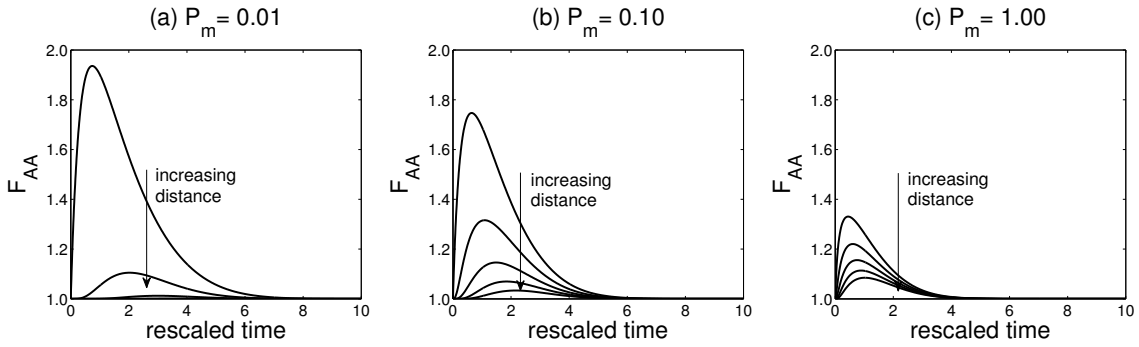


Figure 4.2: Increasing the movement rate, P_m, causes the pair-wise correlation functions, equation (4.3.4), to decay faster. We plot $F_{AA}(s)$ for s increasing from Δ to 3Δ in steps of 0.5Δ. These results are in 2D for $P_p = 0.1$ and $P_d = 0$.

our reaction term is $[1 - c_A F(\Delta)]$. To determine the sign of this, we need to determine the magnitude of $c_A F(\Delta)$. Remembering our definition of the correlation function, we can rewrite this as $\rho^{(2)}(A_l, A_m)/c_A$ where sites **l** and **m** are a distance Δ apart. The numerator of this will always be less than or equal to the denominator because of the following conservation law:

$$\rho^{(2)}(A_l, A_m) + \rho^{(2)}(A_l, 0_m) = \rho^{(1)}(A_l) = c_A. \tag{4.6.2}$$

Therefore, $c_A F(\Delta) \leq 1$ and $[1 - c_A F(\Delta)] \geq 0$. Thus our reaction term will never be positive. Therefore an increase in P_p will lead to a decrease in occupancy correlation for that specific distance. At first this seems counterintuitive, as we expect an increase in the proliferation rate to cause a build-up of correlations. However, this can be explained by considering the effect on the 1-point and 2-point distribution functions. A proliferation event at site l will increase the total agent density, c_A. It will not necessarily increase the 2-point distribution function for lattice sites at distance $l > \Delta$ apart. The correlation function is proportional to the 2-point distribution function, and inversely proportional to the squared 1-point distribution function. Therefore, we expect the correlation function at this specific distance to decrease, as the 1-point distribution function will definitely increase. We also note that at the boundary of

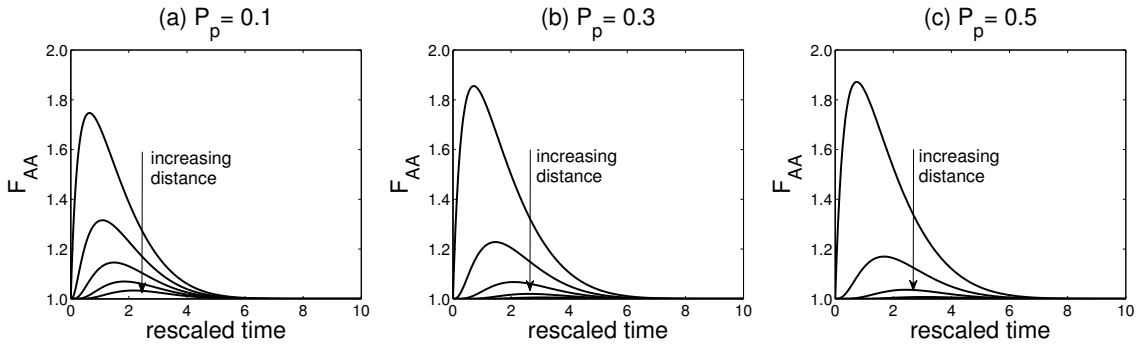


Figure 4.3: Increasing the proliferation rate, P_p, causes an increase in the pair-wise correlation functions at the boundary ($s = \Delta$) but a decrease in the correlation functions at larger distances, $s > \Delta$. The distance, s, increases from Δ to 3Δ in steps of 0.5Δ. Results shown are in 2D for $P_m = 0.1$ and $P_d = 0$.

our correlations PDE, in the expression for the evolution of $F(\Delta)$, equation (4.3.5), proliferation can have an overall positive effect on the build-up of correlations. We can think of the boundary condition in the PDE as acting as a source of correlations which then become dispersed by movement. This phenomenon occurs as we only consider nearest-neighbour proliferation mechanisms. In Figure 4.3, we see that increasing P_p, whilst keeping P_m constant, leads to an increase in the correlation function at the boundary ($s = \Delta$) but a decrease in the correlation function at all other distances.

Using a PDE approximation also enables us to examine the behaviour of the correlations as the distance varies at a specific time, as shown in Figure 4.4. As we expect, the correlation function decreases as the distance increases, reaching a steady-state of unity, confirming equation (4.3.6). We see, especially at later times, that the correlation function does not decay to unity until approximately 10Δ which would require a large number of coupled ODEs in the previous formalism. We have also shown the correlation functions from our averaged simulations, for which the agreement with the PDE is excellent. We also examine the PDE over space and time simultaneously, in Figure 4.5, confirming the observations we have already discussed.

We test our PDE approximation for the parameter space described in Chapter 2 and the further parameter investigations outlined in Section 4.7. We find that

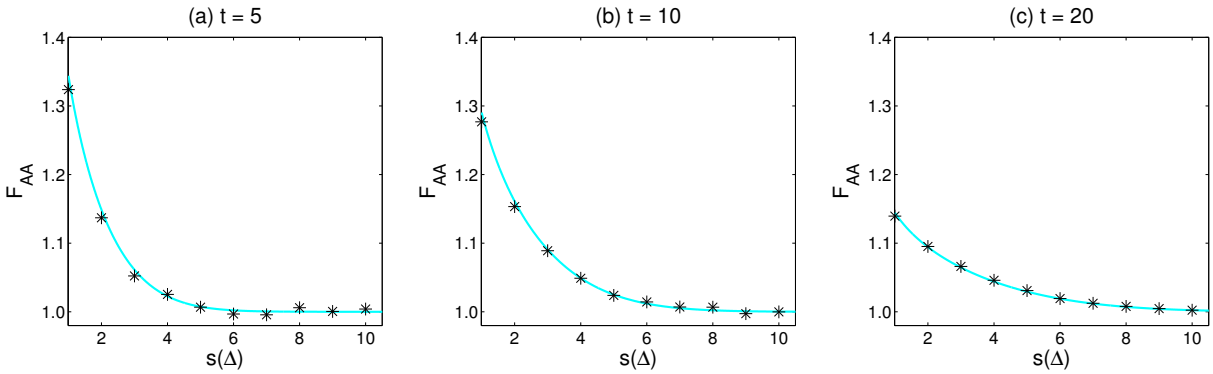
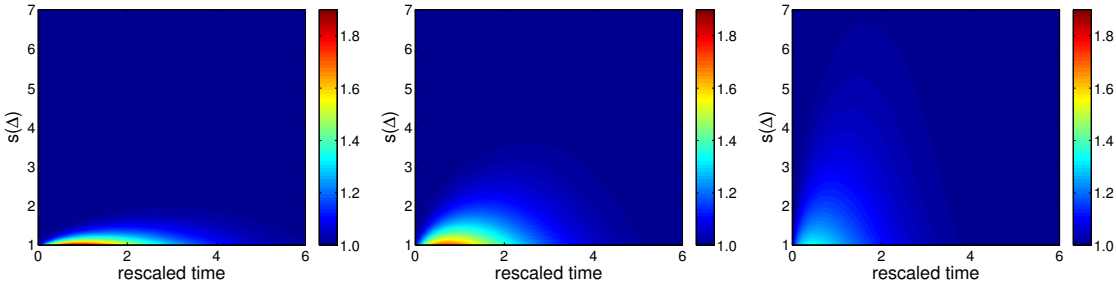


Figure 4.4: The correlation function decreases as the distance between lattice sites increases. Results here are for three different times all with parameters $P_m = 1, P_p = 0.1, P_d = 0.01$ in 2D. The curve in cyan displays results from the solution of the 2D PDE, equation (4.3.4), whilst the black asterisks are from averaged discrete data.



(a) $P_m = 0.1, P_p = 1.0$ (b) $P_m = 1.0, P_p = 1.0$ (c) $P_m = 1.0, P_p = 0.1$

Figure 4.5: The correlation function across space and time for three different parameter sets in 2D, using equation (4.3.4). In all cases $P_d = 0.01$. We see that a lower level of proliferation relative to migration leads to the correlations for small distances having smaller magnitudes, but the long-range correlations being more significant. It also leads to the correlations decaying more rapidly, as agents are moving at a higher rate than they are proliferating thereby enabling the correlations to disperse more quickly.

our approximation performs very well throughout the parameter space in which it is appropriate to close the hierarchy of correlation equations at the pair level to predict the average behaviour of the IBMs.

4.7 Parameter investigations

The correlations model for uniform conditions presented in Chapter 2 was only tested in parameter regions where $P_p/P_m \leq 1$. We observed in Chapter 3, the case with non-uniform initial conditions, that high values of P_p/P_m could lead to the correlations model not predicting the average discrete behaviour accurately. We investigate this in the case with uniform initial conditions, adding to the parameter investigations performed in Chapter 2, by also considering regions of parameter space where motility is reduced relative to proliferation. Thus we keep P_p at unity and vary P_m and P_d accordingly. We focus our attention on 2D and 3D systems as these are more relevant biologically, and the correlations model is more suited to higher dimensions, as seen in Chapter 2 (Figure 2.11). We first look at the case where there is no death ($P_d = 0$) and investigate different rates of movement in 2D and 3D as seen in Figure 4.6. We find that the agreement between the correlations PDE and the averaged IBM improves as the rate of movement increases. We also find that, for the same set of parameters, better agreement between the correlations equations and the averaged IBM is seen in 3D than in 2D, which agrees with the results of Chapter 2. We find that the PDE approximation is always an improvement upon the MFA.

We then introduce death into our system and examine the density results, finding that our correlations PDE still provides good agreement with the averaged IBM. In Figure 4.7, we show results for various movement rates in 2D and 3D, where the death rate is 40% of the proliferation rate. We note that, when death is non-zero, the MFA generally does not reach the same steady-state as the averaged IBM, but that the correlations PDE provides excellent agreement. Thus, for some regions of parameter space, it is the asymptotic, as well as the transient, average discrete density that is not accurately predicted by the MFA.

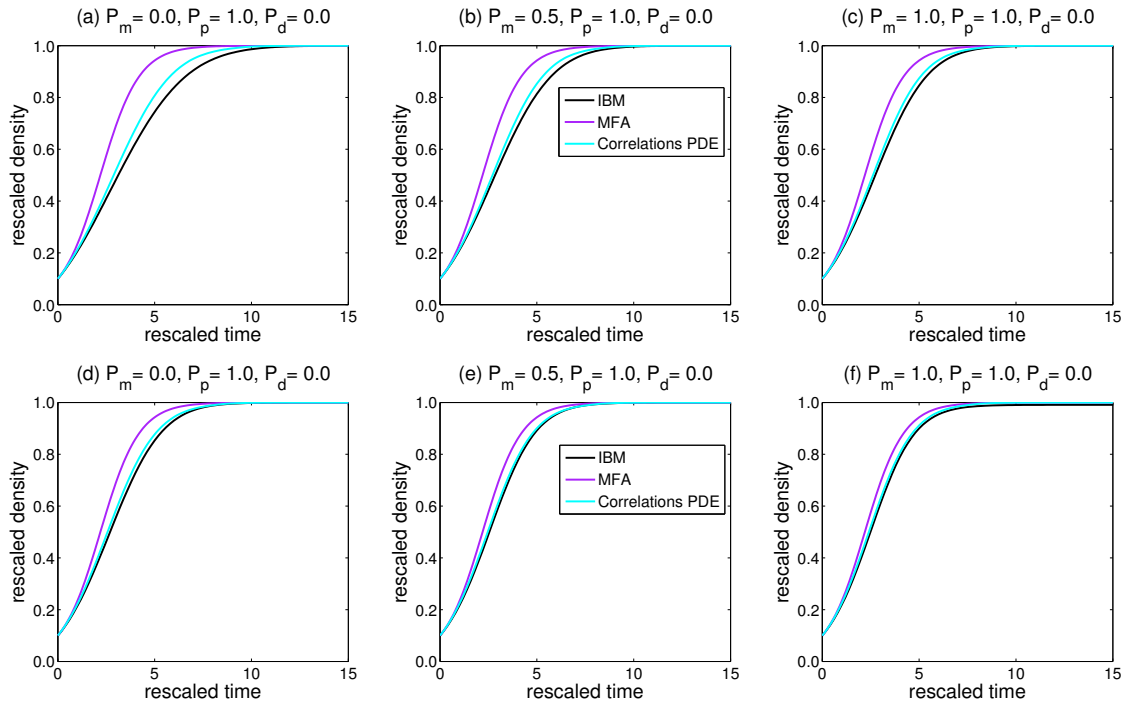


Figure 4.6: We investigate the effects of increasing the movement rate, P_m, in 2D (top row) and 3D (bottom row) when there is no death, $P_d = 0$. An increased movement rate leads to a closer agreement between the correlations PDE (using the equations in Sections 4.3 and 4.4) and the averaged discrete behaviour. The agreement also improves in higher dimensions for the same parameters. The correlations ODEs prediction of the averaged discrete behaviour (not shown) are very similar to, or slightly worse than, the correlations PDE predictions.

4.7.1 Extinction region

In a very small region of parameter space our discrete model quickly tends to extinction and this behaviour is not reflected by either the MFA or the model with correlations included (either the ODE or PDE formulation), although inclusion of the correlations still provides a better approximation than the MFA to the averaged discrete density. This can be seen in Figure 4.8. When $P_m = 0$, as it is for Figure 4.8, our system reduces to a contact process, thus we can turn to theory developed in this area for further insight [Dickman, 1989; Jensen and Dickman, 1993]. The transition from positive to vanishing steady-state densities corresponds to the state phase transition of the contact process [Dickman, 1989; Jensen and Dickman, 1993], which has

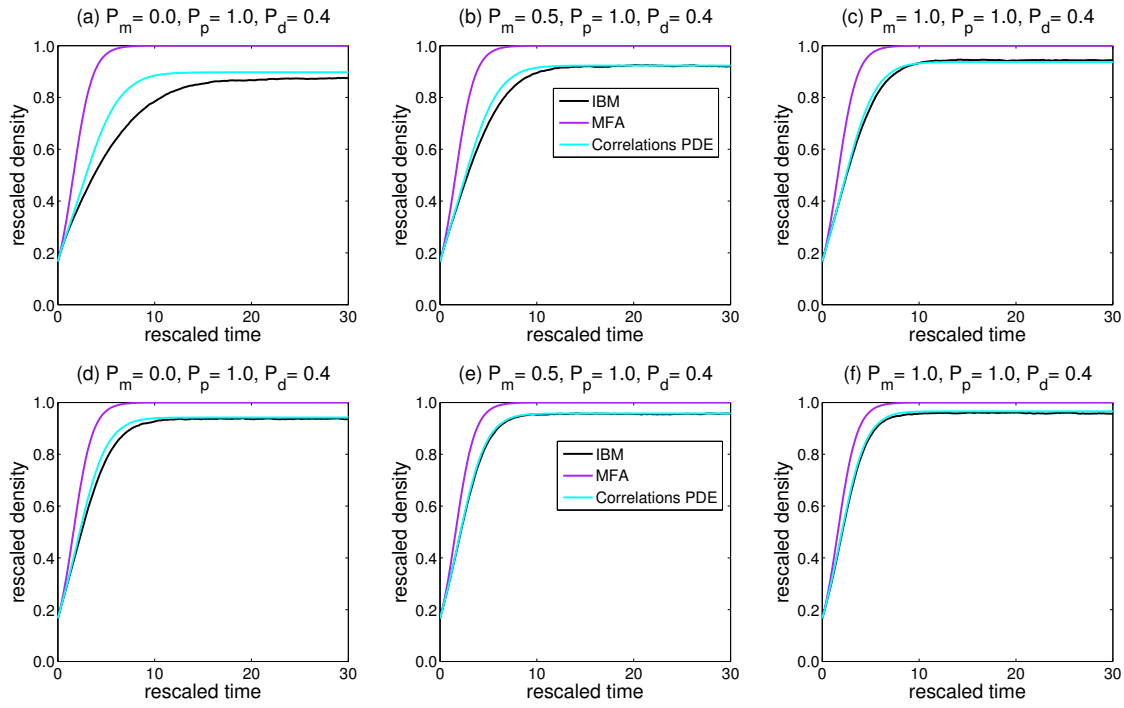


Figure 4.7: We investigate the effects on our approximation of increasing the movement rate, P_m, in 2D (top row) and 3D (bottom row) for a non-zero death rate, $P_d \neq 0$. An increased movement rate leads to a closer agreement between the correlations PDE (using the equations in Sections 4.3 and 4.4) and the averaged discrete behaviour. The correlations ODE predictions of the averaged discrete behaviour (not shown) are very similar to, or slightly worse than, the correlations PDE predictions.

been investigated analytically by series expansions [Jensen and Dickman, 1993]. We note that in these regions of parameter space, the KSA does not perform as well, as discussed by Simpson et al. [2014], thus it is not surprising that our model is not an accurate predictor of the averaged discrete data in these regions of parameter space. Thus, whilst our correlations model generally provides an accurate estimation of the averaged discrete behaviour, we must be cautious when using it in certain regions of parameter space, particularly when there is a high probability of species extinction.

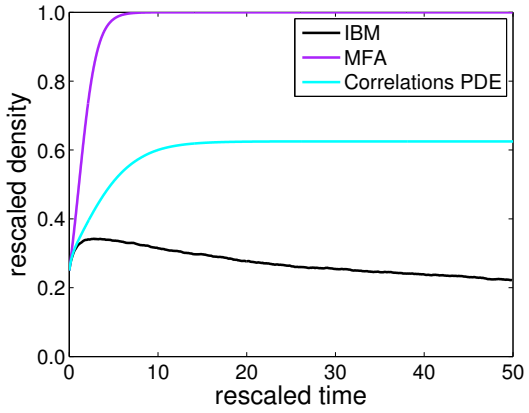


Figure 4.8: For high rates of death, P_d, we see the IBM will eventually lead to population extinction. This behaviour is not reflected by the MFA or the correlations PDE. Results shown here are for the 2D case with $P_m = 0.0, P_p = 1.0, P_d = 0.6$.

4.8 Conclusion

In this chapter, we have provided a tractable method for incorporating spatial correlations into mean-field models. We circumvented the difficulties existing in a previous correlations model [Baker and Simpson, 2010], described in Chapter 2, by simplifying the system of correlation ODEs into a PDE. We demonstrated the success of our method, and investigated the behaviour over a wider range of parameter space than that investigated previously [Baker and Simpson, 2010]. A notable discrepancy between our model and the average discrete behaviour is observed when the death rates are relatively high thus leading to extinction, the absorbing state of the system. In future, we could examine this transition more closely, as done by Jensen and Dickman [1993], as well as develop methods which are able to accurately predict the averaged discrete behaviour in this region of parameter space.

The method discussed thus far in this thesis have only been suitable for systems with a single species. However, many biological systems involve multiple interacting species. Therefore, in the next chapter, we extend the PDE correlations model developed in this chapter to account for multiple species, each with independently prescribed proliferation, migration and death parameters.

Chapter 5

A correlations model for the multi-species case

In the previous chapter, we developed a simplified correlations model for a one-species system undergoing proliferation, movement and death on a regular volume-excluding lattice. This model builds on one from the literature, described in Chapter 2, which incorporates spatial correlations into a continuum model via a system of ODEs [Baker and Simpson, 2010]. The distances in the ODEs for higher dimensions are irregular, thus making it challenging to manually derive them. Expressing this system of irregular ODEs as a PDE simplifies this model, as we demonstrated in the previous chapter, making it more amenable to extensions. In this chapter, we extend the correlations PDE model of the previous chapter to a system with multiple species. We describe the general method for the N-species case, and show some key results for the two-species case.

5.1 Introduction

In this thesis so far we have observed that models including spatial correlations provide a better approximation of the averaged discrete density than do the correspond-

ing MFA models. However, all our models so far have focussed on the one-species case, but biological systems frequently involve multiple species occupying the same environment. For example, it is well known that cells within a tumour tend not to be phenotypically identical; this heterogeneity has led to questioning of the efficacy of many cancer treatments currently in use [Denison and Bae, 2012; Gerlinger et al., 2012; Marusyk et al., 2012]. It has also been observed that the presence of one species can have a detrimental effect on the survival of another, as in the case of red and grey squirrels in Britain [Gurnell, 1987; MacKinnon, 1978; Okubo et al., 1989]. In order to understand and accurately predict the behaviour of multiple species in an environment it is essential that we develop robust mathematical frameworks to describe them.

Much work has been devoted to the development of phenomenological descriptions of multi-species systems in biology, leading to many important advances in a number of areas [Brauer and Castillo-Chavez, 2001; Murray, 2002]. However, these models are frequently based on the MFA, which assumes that there are no spatial correlations between the locations of individuals. As for the one-species case, this assumption is not always valid and thus may lead to inaccurate predictions if used inappropriately [Webb et al., 2007]. Therefore, for situations where the MFA is not a valid model, we would like to develop methods incorporating spatial correlations for systems with multiple species.

As with Chapters 2 and 4, we consider a birth-death-migration process on a volume-excluding lattice with square elements and periodic boundary conditions. However, in this chapter, we now have N species, $N > 1$, each of which has independently prescribed proliferation, migration, and death rates. We show how to derive PDE descriptions of site occupancy correlation functions for this N-species case, highlighting differences from the previous chapter. The species do not have a direct effect on one another, but are indirectly competing for space due to volume-

exclusion. We show the failure of the MFA at predicting the averaged individual-level behaviour in certain situations for the two-species case. This failure could have severe consequences if the MFA is used to model a situation for which it is not an appropriate choice. In contrast, we demonstrate that our correlations PDE model leads to more accurate predictions of the averaged discrete behaviour over a wide range of parameter space.

We begin, in Section 5.2, by describing the general system we are modelling and the details of the mathematical methods used to derive the relevant equations. In Section 5.3, we discuss the two-species case, comparing our numerical results to those from averaged discrete simulations, for systems both with and without agent death. We discuss scenarios whereby the mean-field model has severe shortcomings, and show that they are often eliminated by incorporation of lattice site occupancy correlations. We also discuss regions of parameter space where neither the MFA nor the correlations model accurately predict the averaged discrete results. Finally, in Section 5.4, we conclude and discuss some future directions.

5.2 The general system

We consider a multi-species contact process with diffusion [Cox and Schinazi, 2009; Liggett, 1999], whereby we have a volume-excluding system with N different species of agents undergoing movement, proliferation, and death. The Ith species has a movement rate per unit time of P_m^I, a proliferation rate per unit time of P_p^I, a death rate per unit time of P_d^I, and an initial density of P_i^I. We have a d-dimensional system on a regular, square lattice with a lattice spacing of Δ. The lattice is initially randomly populated, thus lattice site occupancies are initially uncorrelated, and the density is approximately spatially uniform. When an agent moves or divides, it chooses its new location at random from its von Neumann neighbourhood, and the event is aborted if the chosen site is already occupied.

5.2.1 Equations for the N-species case

As with the one-species case, we make use of the k-point distribution functions, $\rho^{(k)}$, which can be thought of as the probability of k-tuplets of sites having specific occupancies [Baker and Simpson, 2010]. In keeping with our work thus far, we close at $\rho^{(3)}$ using the KSA [Kirkwood, 1935; Kirkwood and Boggs, 1942] which expresses $\rho^{(3)}$ in terms of lower order distribution functions (equation (2.5.14)). We again use correlation functions, as defined in equation (2.5.1), to determine the correlation between lattice site occupancies a given distance apart. As we have spatially uniform initial conditions, the correlation functions depend only on the distance between the lattice sites.

In the single-species model with uniform initial conditions, we have an ODE describing the evolution of the agent density, coupled to a system of ODEs for the correlation functions [Baker and Simpson, 2010]. This system of ODEs can instead be written as a PDE with a boundary ODE for correlations at a distance of Δ and a far-field boundary condition, as we demonstrated in the previous chapter. With multiple species, we will obviously have a larger system of equations. We now discuss how to determine the number of equations we will be required to solve for a system with N species.

We begin with the equations for evolution of the agent densities. Similar to the one-species case, the homogeneous initial conditions allow us to express the 1-point distribution function for the Ith species by the normalized concentration of the Ith species:

$$\rho^{(1)}(I_l) = c_I, \tag{5.2.1}$$

where c_I is given by

$$c_I(t) = \frac{1}{\eta_{tot}} \sum_i \mathbb{1}_I\{\sigma_i\}, \tag{5.2.2}$$

with η_{tot} being the total number of lattice sites. The sites must be occupied by one of the species, or vacant, thus leading to the following conservation statement

$$\sum_{J=0}^{N} c_J = 1, \tag{5.2.3}$$

where c_0 is the normalized density of empty sites. Thus we can reduce the number of unknown densities from $N+1$ to N using the conservation statement to write c_0 as a function of the other densities:

$$c_0 = 1 - \sum_{J=1}^{N} c_J. \tag{5.2.4}$$

We use the evolution of the 2-point functions to derive the evolution of our correlation functions, according to the definition in equation (2.5.1). The conservation statements for the 2-point functions take the form

$$\sum_{J=0}^{N} \rho^{(2)}(I_l, J_m) = \rho^{(1)}(I_l) = c_I. \tag{5.2.5}$$

In total, we have $N+1$ second order conservation statements. We use these to reduce the number of unknown correlation functions, choosing to eliminate all F_{I0} where $I = 0, 1, 2, ..., N$ using expressions of the following form

$$F_{I0}(|\mathbf{l}-\mathbf{m}|) = \frac{1 - \sum_{J=1}^{N} c_J F_{IJ}(|\mathbf{l}-\mathbf{m}|)}{1 - \sum_{J=1}^{N} c_J}. \tag{5.2.6}$$

Therefore, in total, we have N unknown densities and $N(N+1)/2$ unknown types of correlation functions. Using PDE descriptions of the correlation functions, as in Chapter 4, leads to $N(N+1)/2$ PDEs, each with a boundary ODE for correlations at a distance of Δ and a far-field boundary condition.

5.2.2 Deriving the MFA

We derive the models in a similar fashion to the single-species case, by using master equations for the evolution of k-point distribution functions. First, we use the 1-point distribution functions to obtain equations for the evolution of the normalized density of each species:

$$\begin{aligned}\frac{\mathrm{d}\rho^{(1)}(I_l)}{\mathrm{d}t} = &\underbrace{P_m^I \sum_n \frac{\alpha_{n,l}}{z}[\rho^{(2)}(0_l, I_n) - \rho^{(2)}(I_l, 0_n)]}_{\text{movement in and out of } l} \\ &+ \underbrace{P_p^I \sum_n \frac{\alpha_{n,l}}{z}\rho^{(2)}(0_l, I_n)}_{\text{proliferation into } l} - \underbrace{P_d^I \rho^{(1)}(I_l)}_{\text{death at } l}.\end{aligned} \tag{5.2.7}$$

Using $\rho^{(1)}(I_l) = c_I$ and the definition of a correlation function, as well as assuming translational invariance, we can simplify this equation to

$$\frac{\mathrm{d}c_I}{\mathrm{d}t} = P_p^I c_I \left[1 - \sum_{J=1}^{N} c_J F_{IJ}(\Delta)\right] - P_d^I c_I. \tag{5.2.8}$$

In the one-species case, the MFA consisted of a single equation. We see, from equation (5.2.8), that in the multi-species case we have equations of a similar form. However, there is now a system of N equations that are coupled by the carrying capacity term. The proliferation term now depends on the densities of the other species and the nearest-neighbour correlation functions. Again, the death term is independent of the correlation functions as a death event does not depend on the occupancies of any other lattice sites in our system. If lattice sites are independent (all $F_{IJ}(|\mathbf{l} - \mathbf{m}|) = 1$), then these equations reduce to the MFA

$$\frac{\mathrm{d}c_I}{\mathrm{d}t} = P_p^I c_I \left[1 - \sum_{J=1}^{N} c_J\right] - P_d^I c_I. \tag{5.2.9}$$

Unsurprisingly, we have recovered a system of coupled logistic equations for our MFA.

5.2.3 The correlations approach

We now turn to the correlations model, in which we are no longer making the assumption of lattice site occupancy independence. We begin with the derivation of the ODE formulation of the correlation functions, similar to the equations described in Chapter 2. We then use the same methods described in Chapter 4 to derive PDE representations for the systems of correlation ODEs.

5.2.3.1 Deriving the correlations ODEs

The correlation functions will take two different forms depending on whether we are considering the spatial correlations between two agents of the same or different species separated by a radial distance of $|\mathbf{l}-\mathbf{m}|$. This differs from the one-species case where we only have one type of correlation function because we only have one agent type. We first consider the evolution of the auto-correlation functions, which apply to agents of the same species at a distance of $|\mathbf{l}-\mathbf{m}|$, by deriving the 2-point distribution functions for the same species case:

$$\begin{aligned}\frac{\mathrm{d}\rho^{(2)}(I_l,I_m)}{\mathrm{d}t} = P_m^I &\left[\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\rho^{(3)}(I_l,0_m,I_n)-\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\rho^{(3)}(I_l,I_m,0_n)\right.\\ &\left.+\sum_{n\neq m}\frac{\alpha_{n,l}}{z}\rho^{(3)}(0_l,I_m,I_n)-\sum_{n\neq m}\frac{\alpha_{n,l}}{z}\rho^{(3)}(I_l,I_m,0_n)\right]\\ &+P_p^I\left[\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\rho^{(3)}(I_l,0_m,I_n)+\sum_{n\neq m}\frac{\alpha_{n,l}}{z}\rho^{(3)}(0_l,I_m,I_n)\right]\\ &+P_p^I\frac{\alpha_{l,m}}{z}\left[\rho^{(2)}(I_l,0_m)+\rho^{(2)}(0_l,I_m)\right]-2P_d^I\rho^{(2)}(I_l,I_m).\end{aligned} \tag{5.2.10}$$

Assuming translational invariance, this simplifies to

$$\begin{aligned}\frac{\mathrm{d}\rho^{(2)}(I_l, I_m)}{\mathrm{d}t} =& 2P_m^I \sum_{n\neq l} \frac{\alpha_{n,m}}{z} \left[\rho^{(3)}(I_l, 0_m, I_n) - \rho^{(3)}(I_l, I_m, 0_n)\right] \\ &+ 2P_p^I \sum_{n\neq l} \frac{\alpha_{n,m}}{z} \rho^{(3)}(I_l, 0_m, I_n) \\ &+ 2P_p^I \frac{\alpha_{l,m}}{z} \rho^{(2)}(I_l, 0_m) - 2P_d^I \rho^{(2)}(I_l, I_m).\end{aligned} \tag{5.2.11}$$

Recall, from equation (2.5.1), that we can also write

$$\frac{\mathrm{d}\rho^{(2)}(I_l, I_m)}{\mathrm{d}t} = c_I^2 \frac{\mathrm{d}F_{II}(|\mathbf{l}-\mathbf{m}|)}{\mathrm{d}t} + 2c_I \frac{\mathrm{d}c_I}{\mathrm{d}t} F_{II}(|\mathbf{l}-\mathbf{m}|). \tag{5.2.12}$$

Therefore the evolution of our auto-correlation functions is given by

$$\begin{aligned}\frac{\mathrm{d}F_{II}(|\mathbf{l}-\mathbf{m}|)}{\mathrm{d}t} =& \frac{2P_m^I}{c_I^2} \sum_{n\neq l} \frac{\alpha_{n,m}}{z} \left[\rho^{(3)}(I_l, 0_m, I_n) - \rho^{(3)}(I_l, I_m, 0_n)\right] \\ &+ \frac{2P_p^I}{c_I^2} \sum_{n\neq l} \frac{\alpha_{n,m}}{z} \rho^{(3)}(I_l, 0_m, I_n) \\ &+ \frac{2P_p^I}{c_I} \frac{\alpha_{l,m}}{z} \lambda_I(|\mathbf{l}-\mathbf{m}|) - 2P_p^I F_{II}(|\mathbf{l}-\mathbf{m}|)\lambda_I(\Delta),\end{aligned} \tag{5.2.13}$$

where

$$\lambda_I(s) = 1 - \sum_{J=1}^{N} c_J F_{IJ}(s), \tag{5.2.14}$$

with $s = |\mathbf{l}-\mathbf{m}|$ being the radial distance between sites l and m. At this point we close our system using the KSA, as defined in equation (2.5.14). A key difference arises at this point between the one-species and N-species problems. In the case with one species, the 3-point distribution functions in the movement terms were eliminated using conservation statements of the following form:

$$\rho^{(2)}(I_l, K_m) = \sum_{J=0}^{N} \rho^{(3)}(I_l, K_m, J_n). \tag{5.2.15}$$

However, in the N-species case, using these conservation equations will no longer eliminate all the $\rho^{(3)}$ terms from the movement expressions. This is due to the ad-

ditional terms in the conservation equations that arise from having more than one species. For example, in the one-species case, the movement term has the following form:

$$\rho^{(3)}(A_l, 0_m, A_n) - \rho^{(3)}(A_l, A_m, 0_n). \tag{5.2.16}$$

We use the following two conservation equations:

$$\rho^{(2)}(A_l, A_m) = \rho^{(3)}(A_l, A_m, A_n) + \rho^{(3)}(A_l, A_m, 0_n), \tag{5.2.17}$$

and

$$\rho^{(2)}(A_l, A_n) = \rho^{(3)}(A_l, A_m, A_n) + \rho^{(3)}(A_l, 0_m, A_n), \tag{5.2.18}$$

to write

$$\begin{aligned}\rho^{(3)}(A_l, 0_m, A_n) - \rho^{(3)}(A_l, A_m, 0_n) = &\rho^{(2)}(A_l, A_n) - \rho^{(3)}(A_l, A_m, A_n) \\ &- \rho^{(2)}(A_l, A_m) + \rho^{(3)}(A_l, A_m, A_n),\end{aligned} \tag{5.2.19}$$

which simplifies to

$$\rho^{(3)}(A_l, 0_m, A_n) - \rho^{(3)}(A_l, A_m, 0_n) = \rho^{(2)}(A_l, A_n) - \rho^{(2)}(A_l, A_m). \tag{5.2.20}$$

Thus, we see that the conservation equations can be used to eliminate the $\rho^{(3)}$ terms from our movement expressions in the one-species case. In the two species case (species A and B), the movement term for species A has the same form as the one-species case. We write the equivalents of equations (5.2.17) and (5.2.18):

$$\rho^{(2)}(A_l, A_m) = \rho^{(3)}(A_l, A_m, A_n) + \rho^{(3)}(A_l, A_m, 0_n) + \rho^{(3)}(A_l, A_m, B_n), \tag{5.2.21}$$

and

$$\rho^{(2)}(A_l, A_n) = \rho^{(3)}(A_l, A_m, A_n) + \rho^{(3)}(A_l, 0_m, A_n) + \rho^{(3)}(A_l, B_m, 0_n). \tag{5.2.22}$$

Using these conservation equations, the movement term for the two species case becomes

$$\begin{aligned}\rho^{(3)}(A_l, 0_m, A_n) - \rho^{(3)}(A_l, A_m, 0_n) = &\rho^{(2)}(A_l, A_n) - \rho^{(3)}(A_l, A_m, A_n) \\ &- \rho^{(3)}(A_l, B_m, A_n) - \rho^{(2)}(A_l, A_m) \\ &+ \rho^{(3)}(A_l, A_m, A_n) + \rho^{(3)}(A_l, A_m, B_n),\end{aligned} \tag{5.2.23}$$

which simplifies to

$$\begin{aligned}\rho^{(3)}(A_l, 0_m, A_n) - \rho^{(3)}(A_l, A_m, 0_n) = &\rho^{(2)}(A_l, A_n) - \rho^{(2)}(A_l, A_m) \\ &+ \rho^{(3)}(A_l, A_m, B_n) - \rho^{(3)}(A_l, B_m, A_n).\end{aligned} \tag{5.2.24}$$

We see that it is no longer possible to eliminate all the $\rho^{(3)}$ terms from our movement expressions. Additionally, when the $\rho^{(3)}$ terms in the conservation equations are expressed using the KSA, the conservation equations no longer hold exactly, as we are using an approximation for each $\rho^{(3)}$ term. This leads to additional discrepancies in our estimation of the averaged discrete behaviour. We can see this in Figure 5.1, where we calculate $\rho^{(2)}(A_l, A_m)$ in three ways:

1. directly from the discrete data;

2. calculating $\rho^{(3)}(A_l, A_m, A_n)$ and $\rho^{(3)}(A_l, A_m, 0_n)$ directly from the discrete data and then expressing $\rho^{(2)}(A_l, A_m)$ using equation (5.2.17);

3. calculating $\rho^{(3)}(A_l, A_m, A_n)$ and $\rho^{(3)}(A_l, A_m, 0_n)$ using the KSA and then expressing $\rho^{(2)}(A_l, A_m)$ using equation (5.2.17).

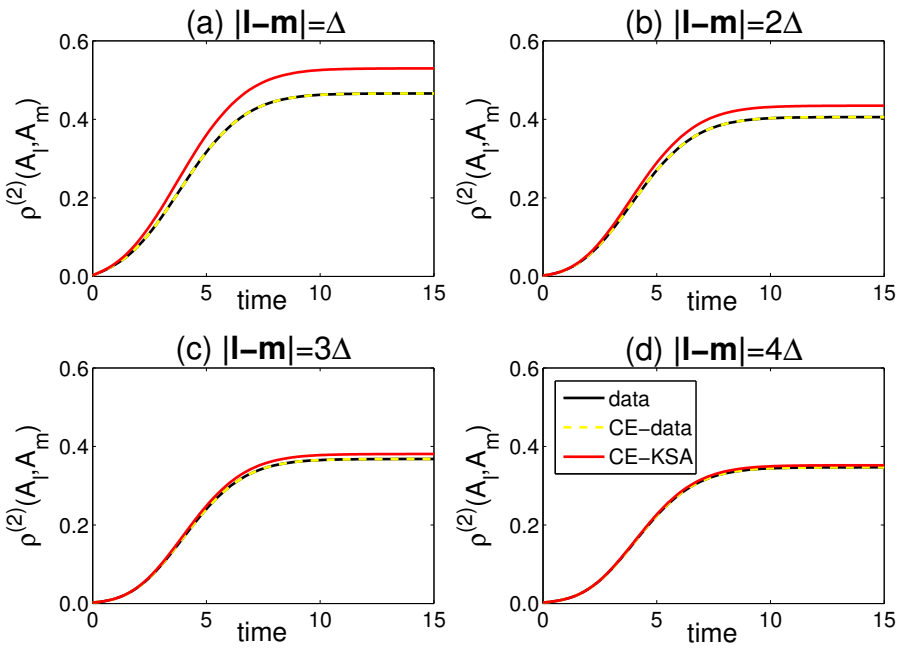


Figure 5.1: Using the KSA on our conservation expressions leads to some discrepancies. Here we show $\rho^{(2)}(A_l, A_m)$ calculated directly from averaged data (black). We also calculate the $\rho^{(3)}$ terms in the conservation equation (equation 5.2.17) directly from the data, confirming that the conservation equation holds (yellow). Additionally, we express each $\rho^{(3)}$ term in the conservation equation using the KSA and discover that the conservation equation no longer holds exactly as we are now approximating the $\rho^{(3)}$ terms (red). We look at lattice distances of Δ to 4Δ increasing in steps of Δ, (a)-(d), and find that the approximation performs better as the distance between lattice sites increases, which agrees with our observations on the KSA in Chapter 2.

We see that using the KSA on each term in the conservation equation leads to some discrepancies from the averaged discrete results, particular when $|\mathbf{l} - \mathbf{m}|$ is small. We remember, from Chapter 2, that the KSA does not perform as well when the lattice sites in question are close together. This explains why using the KSA on the $\rho^{(3)}$ terms in the conservation equation does not give us very good agreement with the discrete data when $|\mathbf{l} - \mathbf{m}|$ is small. Thus we do not rewrite our movement terms using the conservation equations, and instead use the KSA directly on the $\rho^{(3)}$ terms in equation (5.2.13). This leads to the movement terms of our equations having a

different form to the one species case. Closing with the KSA leads to

$$\begin{aligned}\frac{\mathrm{d}F_{II}(|\mathbf{l}-\mathbf{m}|)}{\mathrm{d}t} =& \frac{2P_m^I}{1-\sum_{J=1}^N c_J}\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\lambda_I(\Delta)\Big[F_{II}(|\mathbf{l}-\mathbf{n}|)-F_{II}(|\mathbf{l}-\mathbf{m}|)\\ &-F_{II}(|\mathbf{l}-\mathbf{n}|)\sum_{J=1}^N c_J F_{IJ}(|\mathbf{l}-\mathbf{m}|)+F_{II}(|\mathbf{l}-\mathbf{m}|)\sum_{J=1}^N c_J F_{IJ}(|\mathbf{l}-\mathbf{n}|)\Big]\\ &+\frac{2P_p^I}{1-\sum_{J=1}^N c_J}\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\lambda_I(|\mathbf{l}-\mathbf{m}|)\lambda_I(\Delta)F_{II}(|\mathbf{l}-\mathbf{n}|)\\ &+\frac{2P_p^I}{c_I}\frac{\alpha_{l,m}}{z}\lambda_I(\Delta)-2P_p^I\lambda_I(\Delta)F_{II}(|\mathbf{l}-\mathbf{m}|).\end{aligned} \tag{5.2.25}$$

We now consider the evolution of the cross-correlation equations, which are applicable to agents of different species separated by a distance of $|\mathbf{l}-\mathbf{m}|$. We again start with the evolution of the 2-point distribution function, which we can once more simplify due to translational invariance, to arrive at

$$\begin{aligned}\frac{\mathrm{d}\rho^{(2)}(I_l,K_m)}{\mathrm{d}t} =& P_m^I\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\left[\rho^{(3)}(K_l,0_m,I_n)-\rho^{(3)}(K_l,I_m,0_n)\right]\\ &+P_m^K\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\left[\rho^{(3)}(I_l,0_m,K_n)-\rho^{(3)}(I_l,K_m,0_n)\right]\\ &+P_p^I\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\rho^{(3)}(K_l,0_m,I_n)+P_p^K\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\rho^{(3)}(I_l,0_m,K_n)\\ &-P_d^I\rho^{(2)}(I_l,K_m)-P_d^K\rho^{(2)}(I_l,K_m).\end{aligned} \tag{5.2.26}$$

From the definition of the correlation function, we obtain

$$\frac{\mathrm{d}\rho^{(2)}(I_l,K_m)}{\mathrm{d}t} = c_Ic_K\frac{\mathrm{d}F_{IK}(|\mathbf{l}-\mathbf{m}|)}{\mathrm{d}t}+\left(c_I\frac{\mathrm{d}c_K}{\mathrm{d}t}+c_K\frac{\mathrm{d}c_I}{\mathrm{d}t}\right)F_{IK}(|\mathbf{l}-\mathbf{m}|). \tag{5.2.27}$$

We relate equations (5.2.26) and (5.2.27) in the same way as before to obtain

$$\begin{aligned}\frac{\mathrm{d}F_{IK}(|\mathbf{l}-\mathbf{m}|)}{\mathrm{d}t} =& \frac{P_m^I}{c_I c_K}\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\left[\rho^{(3)}(K_l,0_m,I_n)-\rho^{(3)}(K_l,I_m,0_n)\right]\\ &+\frac{P_m^K}{c_I c_K}\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\left[\rho^{(3)}(I_l,0_m,K_n)-\rho^{(3)}(I_l,K_m,0_n)\right]\\ &+\frac{P_p^I}{c_I c_K}\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\rho^{(3)}(K_l,0_m,I_n)+\frac{P_p^K}{c_I c_K}\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\rho^{(3)}(I_l,0_m,K_n)\\ &-P_p^I F_{IK}(|\mathbf{l}-\mathbf{m}|)\lambda_I(\Delta)-P_p^K F_{IK}(|\mathbf{l}-\mathbf{m}|)\lambda_K(\Delta).\end{aligned} \tag{5.2.28}$$

Again, we use the KSA to close our equations, without rewriting our movement terms using the conservation equations, obtaining

$$\begin{aligned}\frac{\mathrm{d}F_{IK}(|\mathbf{l}-\mathbf{m}|)}{\mathrm{d}t} =& \frac{P_m^I}{1-\sum_{J=1}^N c_J}\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\lambda_I(\Delta)\Big[F_{IK}(|\mathbf{l}-\mathbf{n}|)-F_{IK}(|\mathbf{l}-\mathbf{m}|)\\ &-F_{IK}(|\mathbf{l}-\mathbf{n}|)\sum_{J=1}^N c_J F_{KJ}(|\mathbf{l}-\mathbf{m}|)+F_{IK}(|\mathbf{l}-\mathbf{m}|)\sum_{J=1}^N c_J F_{KJ}(|\mathbf{l}-\mathbf{n}|)\Big]\\ &+\frac{P_m^K}{1-\sum_{J=1}^N c_J}\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\lambda_K(\Delta)\Big[F_{IK}(|\mathbf{l}-\mathbf{n}|)-F_{IK}(|\mathbf{l}-\mathbf{m}|)\\ &-F_{IK}(|\mathbf{l}-\mathbf{n}|)\sum_{J=1}^N c_J F_{IJ}(|\mathbf{l}-\mathbf{m}|)+F_{IK}(|\mathbf{l}-\mathbf{m}|)\sum_{J=1}^N c_J F_{IJ}(|\mathbf{l}-\mathbf{n}|)\Big]\\ &+\frac{P_p^I}{1-\sum_{J=1}^N c_J}\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\lambda_K(|\mathbf{l}-\mathbf{m}|)\lambda_I(\Delta)F_{IK}(|\mathbf{l}-\mathbf{n}|)\\ &+\frac{P_p^K}{1-\sum_{J=1}^N c_J}\sum_{n\neq l}\frac{\alpha_{n,m}}{z}\lambda_I(|\mathbf{l}-\mathbf{m}|)\lambda_K(\Delta)F_{IK}(|\mathbf{l}-\mathbf{n}|)\\ &-P_p^I\lambda_I(\Delta)F_{IK}(|\mathbf{l}-\mathbf{m}|)-P_p^K\lambda_K(\Delta)F_{IK}(|\mathbf{l}-\mathbf{m}|).\end{aligned} \tag{5.2.29}$$

5.2.3.2 Deriving the correlations PDEs

As we saw in Chapter 2, encoding the system of ODEs in higher dimensions is not straightforward due to the non-uniform distances. This will be even more challenging with the multi-species case. Thus, following the simplification in Chapter 4, we now derive PDE descriptions of the correlation functions. We demonstrate the method in 2D, although the same principles apply in other dimensions, as seen in Chapter 4. We begin with the PDE for the auto-correlation functions. The PDE will be defined for the spatial domain $\Delta < s < s_{max}$, where s_{max} is the distance beyond which lattice sites are assumed to be independent. As with the single-species case, we expect lattice sites that are far apart to be independent, thus leading to the boundary condition

$$F_{II}(s = s_{max}) = 1. \tag{5.2.30}$$

To obtain the boundary condition at $s = \Delta$ we use equation (5.2.25) which gives

$$\begin{aligned}\frac{\mathrm{d}F_{II}(\Delta)}{\mathrm{d}t} = &\frac{P_m^I}{2(1-\sum_{J=1}^N c_J)}\lambda_I(\Delta)\Bigg[-3F_{II}(\Delta) + \left\{2F_{II}(\sqrt{2}\Delta) + F_{II}(2\Delta)\right\}\lambda_I(\Delta) \\ &+F_{II}(\Delta)\sum_{J=1}^N c_J\left\{2F_{IJ}(\sqrt{2}\Delta) + F_{IJ}(2\Delta)\right\}\Bigg] \\ &+\frac{P_p^I}{2(1-\sum_{J=1}^N c_J)}\lambda_I(\Delta)^2\left[2F_{II}(\sqrt{2}\Delta) + F_{II}(2\Delta)\right] \\ &+\frac{P_p^I}{2c_I}\lambda_I(\Delta) - 2P_p^I\lambda_I(\Delta)F_{II}(\Delta).\end{aligned} \tag{5.2.31}$$

To obtain the PDE, we rewrite equation (5.2.25) using Cartesian coordinates, and then perform a Taylor expansion of the correlation function around (x, y). Finally we take $\Delta \to 0$, explicitly retaining our boundary terms, and return to radial coordinates, obtaining the following 2D PDE:

$$\frac{\partial F_{II}(s)}{\partial t} = \frac{P_m^I}{2(1-\sum_{J=1}^N c_J)}\lambda_I(\Delta)\left[\nabla^2 F_{II}\lambda_I(s) + F_{II}(s)\sum_{J=1}^N c_J \nabla^2 F_{IJ}(s)\right] + \frac{2P_p^I}{1-\sum_{J=1}^N c_J}\lambda_I(\Delta)\lambda_I(s)F_{II}(s) - 2P_p^I\lambda_I(\Delta)F_{II}(s), \qquad \Delta < s < s_{max}, \tag{5.2.32}$$

where ∇^2 is the polar Laplacian with no angular dependence.

We do the same for the cross-correlation functions to obtain the following PDE:

$$\begin{aligned}\frac{\partial F_{IK}(s)}{\partial t} =& \frac{P_m^I}{4(1-\sum_{J=1}^N c_J)}\lambda_I(\Delta)\left[\nabla^2 F_{IK}(s)\lambda_K(s) + F_{IK}(s)\sum_{J=1}^N c_J\nabla^2 F_{KJ}(s)\right] \\ &+\frac{P_m^K}{4(1-\sum_{J=1}^N c_J)}\lambda_K(\Delta)\left[\nabla^2 F_{IK}(s)\lambda_I(s) + F_{IK}(s)\sum_{J=1}^N c_J\nabla^2 F_{IJ}(s)\right] \\ &+ \frac{P_p^I}{1-\sum_{J=1}^N c_J}\lambda_I(\Delta)\lambda_K(s)F_{IK}(s) + \frac{P_p^K}{1-\sum_{J=1}^N c_J}\lambda_K(\Delta)\lambda_I(s)F_{IK}(s) \\ &- P_p^I\lambda_I(\Delta)F_{IK}(s) - P_p^K\lambda_K(\Delta)F_{IK}(s), \qquad \Delta < s < s_{max},\end{aligned} \tag{5.2.33}$$

with boundary conditions

$$\begin{aligned}\frac{\mathrm{d}F_{IK}(\Delta)}{\mathrm{d}t} =& \frac{P_m^I}{4(1-\sum_{J=1}^N c_J)}\lambda_I(\Delta)\Bigg[-3F_{IK}(\Delta) + \left\{2F_{IK}(\sqrt{2}\Delta) + F_{IK}(2\Delta)\right\}\lambda_K(\Delta) \\ &+F_{IK}(\Delta)\sum_{J=1}^N c_J\left\{2F_{KJ}(\sqrt{2}\Delta) + F_{KJ}(2\Delta)\right\}\Bigg] \\ &+ \frac{P_m^K}{4(1-\sum_{J=1}^N c_J)}\lambda_K(\Delta)\Bigg[-3F_{IK}(\Delta) + \left\{2F_{IK}(\sqrt{2}\Delta) + F_{IK}(2\Delta)\right\}\lambda_I(\Delta) \\ &+F_{IK}(\Delta)\sum_{J=1}^N c_J\left\{2F_{IJ}(\sqrt{2}\Delta) + F_{IJ}(2\Delta)\right\}\Bigg] \\ &+ \frac{P_p^I + P_p^K}{4(1-\sum_{J=1}^N c_J)}\lambda_I(\Delta)\lambda_K(\Delta)\left[2F_{IK}(\sqrt{2}\Delta) + F_{IK}(2\Delta)\right] \\ &- P_p^I\lambda_I(\Delta)F_{IK}(\Delta) - P_p^K\lambda_K(\Delta)F_{IK}(\Delta),\end{aligned} \tag{5.2.34}$$

and

$$F_{IK}(s = s_{max}) = 1. \tag{5.2.35}$$

5.3 A specific case: two species

We demonstrate the utility of our method in 2D for the $N = 2$ case, whereby we have two species (A and B). We again compare our mean-field and correlations models to results from averaged discrete simulations.

5.3.1 The discrete system

To obtain the averaged discrete results, we use the Gillespie algorithm presented in Chapter 2, but adapted for two species:

1. set $t = 0$;

2. initialize the lattice by generating a uniform random number, r_i, in the interval [0,1], for each lattice site. If $r_i \leq P_i^A$, an agent of species A is placed at that lattice site. If $P_i^A < r_i \leq P_i^A + P_i^B$, an agent of species B is placed at that lattice site;

3. with $Q_A(t)$ and $Q_B(t)$ being the number of agents of type A and B, respectively, calculate the total propensity function, $a_0 = (P_m^A + P_p^A + P_d^A)Q_A(t) + (P_m^B + P_p^B + P_d^B)Q_B(t)$;

4. calculate the time, τ, to the next event using $\tau = (1/a_0)\log(1/r_1)$, where r_1 is a uniform random number in the interval [0,1];

5. decide which event occurs by calculating $R = a_0 r_2$, where r_2 is another uniform random number in the interval [0,1]. R is used to deduce what event occurs according to the following:

- if $R \in [0, P_m^A Q_A(t))$, a movement event for species A will occur. An agent of species A is chosen uniformly at random, and one of its neighbours in a von Neumann neighbourhood is also chosen uniformly at random as the target site for a movement event. If the target site is empty, the chosen agent moves to that site, otherwise the event is aborted;
- if $R \in [P_m^A Q_A(t), (P_m^A + P_p^A)Q_A(t))$, a proliferation event for species A will occur. An agent of species A is chosen uniformly at random, and one of its neighbours in a von Neumann neighbourhood is also chosen uniformly at random as the target site for a proliferation event. If the target site is empty, a daughter cell is placed in it, otherwise the event is aborted;
- if $R \in [(P_m^A + P_p^A)Q_A(t), (P_m^A + P_p^A + P_d^A)Q_A(t))$, a death event for species A will occur. An agent of species A is chosen uniformly at random and removed from the system;
- if $R \in [(P_m^A + P_p^A + P_d^A)Q_A(t), (P_m^A + P_p^A + P_d^A)Q_A(t) + P_m^B Q_B(t))$, a movement event for species B will occur. An agent of species B is chosen uniformly at random, and one of its neighbours in a von Neumann neighbourhood is also chosen uniformly at random as the target site for a movement event. If the target site is empty, the chosen agent moves to that site, otherwise the event is aborted;
- if $R \in [(P_m^A + P_p^A + P_d^A)Q_A(t) + P_m^B Q_B(t), (P_m^A + P_p^A + P_d^A)Q_A(t) + (P_m^B + P_p^B)Q_B(t))$, a proliferation event for species B will occur. An agent of species B is chosen uniformly at random, and one of its neighbours in a von Neumann neighbourhood is also chosen uniformly at random as the target site for a proliferation event. If the target site is empty, a daughter cell is placed in it, otherwise the event is aborted;

- otherwise, a death event for species B will occur. An agent of species B is chosen uniformly at random and removed from the system;

6. update $Q_A(t)$ and $Q_B(t)$ depending on what event, if any, occurred;

7. update time by setting $t \to t + \tau$;

8. repeat from step 3 until the stipulated final time is reached.

We average over 100 realisations, for a 100×100 lattice. For all results presented in this chapter, we use identical initial conditions, $c_A(0) = c_B(0) = 0.05$. We test our discrete results against the MFA in the high migration limit, whereby we know the MFA holds thus they should be identical. We confirm that they do give identical results in this parameter regime. Additionally, we allow species A and B to be identical and confirm that our results are identical to those in the one-species case.

5.3.2 The MFA

The MFA equations for the two species case, from equation (5.2.9), are:

$$\frac{\mathrm{d}c_A}{\mathrm{d}t} = P_p^A c_A(1 - c_A - c_B) - P_d^A c_A, \tag{5.3.1}$$

and

$$\frac{\mathrm{d}c_B}{\mathrm{d}t} = P_p^B c_B(1 - c_A - c_B) - P_d^B c_B. \tag{5.3.2}$$

We solve the system of coupled logistic ODEs for the MFA using a fourth order Runge Kutta method [Press et al., 2007] with a time step, δt, of 0.1 (smaller time steps were tested to confirm the accuracy of this choice). We test the MFA against our discrete case in the high migration limit, where we expect them to be identical due to the system being well-mixed in this parameter regime. We also confirm that the MFA reduces to the one-species case when the two species are identical.

5.3.3 The correlations PDEs

For the correlations PDE model with two species, we have a system with two ODEs (for the evolution of the cell density of each species) and three PDEs (two for the evolution of the auto-correlation functions and one for that of the cross-correlation function), with each PDE having an ODE at the $s = \Delta$ boundary as well as a far-field boundary condition. From the general equation (5.2.8) we can obtain the equations for the evolution of species A and B, which are

$$\frac{\mathrm{d}c_A}{\mathrm{d}t} = P_p^A c_A \left[1 - F_{AA}(\Delta)c_A - F_{AB}(\Delta)c_B\right] - P_d^A c_A, \tag{5.3.3}$$

and

$$\frac{\mathrm{d}c_B}{\mathrm{d}t} = P_p^B c_B \left[1 - F_{AB}(\Delta)c_A - F_{BB}(\Delta)c_B\right] - P_d^B c_B, \tag{5.3.4}$$

respectively. From equation (5.2.25) for the auto-correlation functions, the equation for the evolution of the auto-correlation function for species A, F_{AA}, is

$$\begin{aligned}\frac{\partial F_{AA}}{\partial t} =& \frac{P_m^A}{2(1-c_A-c_B)}\lambda_A(\Delta)\left\{\nabla^2 F_{AA}\lambda_A(s) + F_{AA}(s)\left[c_A F_{AA}(s) + c_B F_{AB}(s)\right]\right\} \\ &+ \frac{2P_p^A}{1-c_A-c_B}\lambda_A(\Delta)\lambda_A(s)F_{AA}(s) - 2P_p^A\lambda_A(\Delta)F_{AA}(s), \qquad \Delta < s < s_{max},\end{aligned} \tag{5.3.5}$$

with boundary conditions

$$\begin{aligned}\frac{\mathrm{d}F_{AA}(\Delta)}{\mathrm{d}t} =& \frac{P_m^A}{2(1-c_A-c_B)}\lambda_A(\Delta)\Big[-3F_{AA}(\Delta) + \left\{2F_{AA}(\sqrt{2}\Delta) + F_{AA}(2\Delta)\right\}\lambda_A(\Delta) \\ &+F_{AA}(\Delta)c_A\left\{2F_{AA}(\sqrt{2}\Delta) + F_{AA}(2\Delta)\right\} \\ &+F_{AA}(\Delta)c_B\left\{2F_{AB}(\sqrt{2}\Delta) + F_{AB}(2\Delta)\right\}\Big] \\ &+ \frac{P_p^A}{2(1-c_A-c_B)}\lambda_A(\Delta)^2\left[2F_{AA}(\sqrt{2}\Delta) + F_{AA}(2\Delta)\right] \\ &+ \frac{P_p^A}{2c_A}\lambda_A(\Delta) - 2P_p^A\lambda_A(\Delta)F_{AA}(\Delta),\end{aligned} \tag{5.3.6}$$

and

$$F_{AA}(s = s_{max}) = 1. \tag{5.3.7}$$

Similarly, the evolution of the auto-correlation function for species B, F_{BB}, is

$$\begin{aligned}\frac{\partial F_{BB}}{\partial t} =& \frac{P_m^B}{2(1-c_A-c_B)}\lambda_B(\Delta)\left\{\nabla^2 F_{BB}\lambda_B(s) + F_{BB}(s)\left[c_B F_{BB}(s) + c_A F_{AB}(s)\right]\right\} \\ &+ \frac{2P_p^B}{1-c_A-c_B}\lambda_B(\Delta)\lambda_B(s)F_{BB}(s) - 2P_p^B\lambda_B(\Delta)F_{BB}(s), \qquad \Delta < s < s_{max},\end{aligned} \tag{5.3.8}$$

with boundary conditions

$$\begin{aligned}\frac{\mathrm{d}F_{BB}(\Delta)}{\mathrm{d}t} =& \frac{P_m^B}{2(1-c_A-c_B)}\lambda_B(\Delta)\Big[-3F_{BB}(\Delta) + \left\{2F_{BB}(\sqrt{2}\Delta) + F_{BB}(2\Delta)\right\}\lambda_B(\Delta) \\ &+F_{BB}(\Delta)c_B\left\{2F_{BB}(\sqrt{2}\Delta) + F_{BB}(2\Delta)\right\} \\ &+F_{BB}(\Delta)c_A\left\{2F_{AB}(\sqrt{2}\Delta) + F_{AB}(2\Delta)\right\}\Big] \\ &+ \frac{P_p^B}{2(1-c_A-c_B)}\lambda_B(\Delta)^2\left[2F_{BB}(\sqrt{2}\Delta) + F_{BB}(2\Delta)\right] \\ &+ \frac{P_p^B}{2c_B}\lambda_B(\Delta) - 2P_p^B\lambda_B(\Delta)F_{BB}(\Delta),\end{aligned} \tag{5.3.9}$$

and

$$F_{BB}(s = s_{max}) = 1. \tag{5.3.10}$$

Finally, the evolution of the cross-correlation function, F_{AB}, from the general equation (equation (5.2.34)), is given by

$$\begin{aligned}\frac{\partial F_{AB}(s)}{\partial t} =& \frac{P_m^A}{4(1-c_A-c_B)}\lambda_A(\Delta)\left\{\nabla^2 F_{AB}(s)\lambda_B(s) + F_{AB}(s)\left[c_A\nabla^2 F_{AB}(s) + c_B\nabla^2 F_{BB}(s)\right]\right\} \\ &+\frac{P_m^B}{4(1-c_A-c_B)}\lambda_B(\Delta)\left\{\nabla^2 F_{AB}(s)\lambda_A(s) + F_{AB}(s)\left[c_A\nabla^2 F_{AA}(s) + c_B\nabla^2 F_{AB}(s)\right]\right\} \\ &+ \frac{P_p^A}{1-c_A-c_B}\lambda_A(\Delta)\lambda_B(s)F_{AB}(s) + \frac{P_p^B}{1-c_A-c_B}\lambda_B(\Delta)\lambda_A(s)F_{AB}(s) \\ &- P_p^A\lambda_A(\Delta)F_{AB}(s) - P_p^B\lambda_B(\Delta)F_{AB}(s),\end{aligned} \tag{5.3.11}$$

with boundary conditions

$$\begin{aligned}\frac{\mathrm{d}F_{AB}(\Delta)}{\mathrm{d}t} =& \frac{P_m^A}{4(1-c_A-c_B)}\lambda_A(\Delta)\Big[+F_{AB}(\Delta)c_A\Big\{2F_{AB}(\sqrt{2}\Delta)+F_{AB}(2\Delta)\Big\}\\ &-3F_{AB}(\Delta)+\Big\{2F_{AB}(\sqrt{2}\Delta)+F_{AB}(2\Delta)\Big\}\lambda_B(\Delta)\\ &+F_{AB}c_B\Big\{2F_{BB}(\sqrt{2}\Delta)+F_{BB}(2\Delta)\Big\}\Big]\\ &\frac{P_m^B}{4(1-c_A-c_B)}\lambda_B(\Delta)\Big[+F_{AB}(\Delta)c_B\Big\{2F_{AB}(\sqrt{2}\Delta)+F_{AB}(2\Delta)\Big\}\\ &-3F_{AB}(\Delta)+\Big\{2F_{AB}(\sqrt{2}\Delta)+F_{AB}(2\Delta)\Big\}\lambda_A(\Delta)\\ &+F_{AB}c_A\Big\{2F_{AA}(\sqrt{2}\Delta)+F_{AA}(2\Delta)\Big\}\Big]\\ &+\frac{P_p^A+P_p^B}{4(1-c_A-c_B)}\lambda_A(\Delta)\lambda_B(\Delta)\Big[2F_{AB}(\sqrt{2}\Delta)+F_{AB}(2\Delta)\Big]\\ &-P_p^A\lambda_A(\Delta)F_{AB}(\Delta)-P_p^B\lambda_B(\Delta)F_{AB}(\Delta),\end{aligned} \tag{5.3.12}$$

and

$$F_{AB}(s=s_{max})=1. \tag{5.3.13}$$

We discretise our equations in time using the backwards Euler method, and we solve the resulting system of non-linear equations using the tridiagonal matrix algorithm [Press et al., 2007] with Picard iteration [Zheng and Bennett, 2002]. We choose $s_{max}=20$, testing $s_{max}=21$ to confirm our results are visually indistinguishable. Again, we test that our model agrees with the averaged discrete behaviour and the MFA in the high migration limit, and confirm it gives the same results as the one-species case if the species are identical.

5.3.4 Systems undergoing proliferation and movement

We first investigate systems undergoing proliferation and movement, but not death ($P_d^A = P_d^B = 0$). Results from a single realisation of the discrete system can be

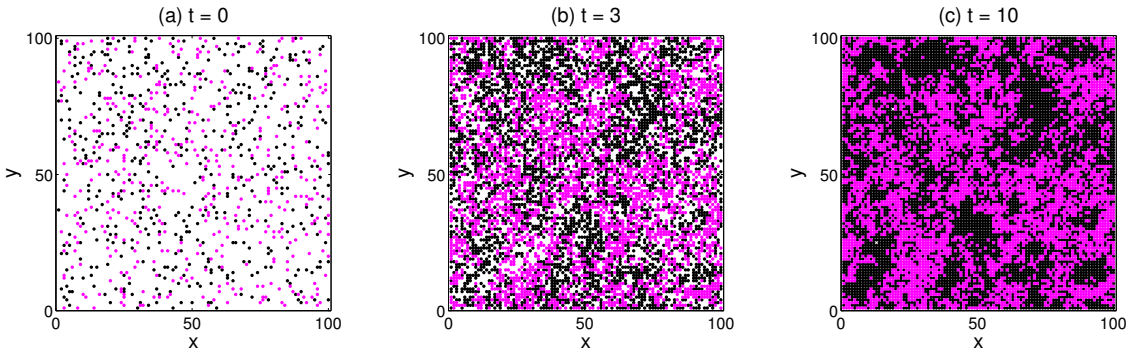


Figure 5.2: An individual discrete simulation showing the build-up of spatial correlations over time. Details of the numerical algorithm can be found in Section 5.3. In this case $P_m^A = 20, P_p^A = 1, P_m^B = 5, P_p^B = 1$. Species A is in black (occupying 56% of the lattice at steady-state) whilst species B is magenta (occupying 44% of the lattice at steady-state).

seen in Figure 5.2. We discuss here three cases where the MFA makes very different predictions from what is observed in the averaged discrete behaviour. In all cases, our correlations PDE method provides excellent predictions of the averaged discrete behaviour, as seen in Figure 5.3.

Firstly, we investigate the case where the two species have the same proliferation rate but different movement rates (Figure 5.3(a)). The MFA predicts that the two species will behave in exactly the same way, leading to each population occupying half of the lattice sites on average. However, the individual realisations show that the species with the higher movement rate will, on average, reach a higher total agent density at the steady-state. Our correlations PDE method provides a good approximation to the averaged individual behaviour.

Secondly, we look at a case whereby the MFA predicts different agent densities at steady-state for the two species, but the averaged discrete data show that we expect the two species to have very similar agent densities at steady-state (Figure 5.3(b)). We find that higher movement rates can compensate for decreased proliferation rates; a feature that is overlooked in the MFA.

The last case we examine is one in which the MFA dominant steady-state species

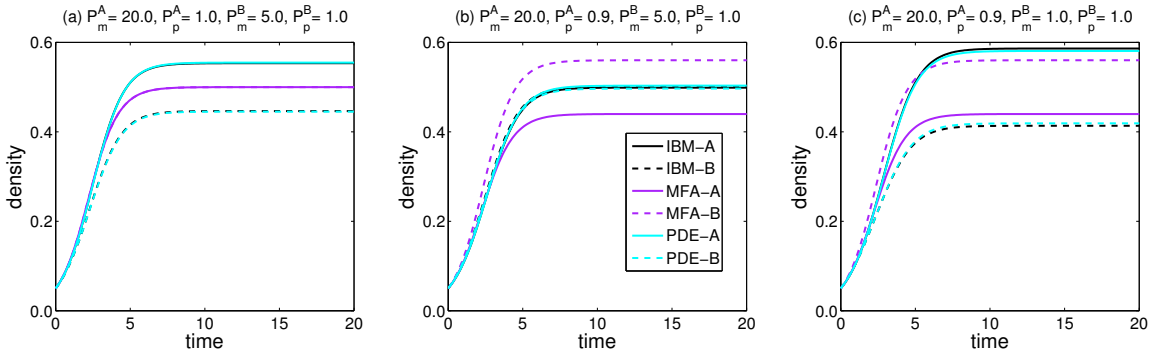


Figure 5.3: Our correlations PDE method (using equations (5.3.3) to (5.3.13)) accurately predicts the averaged discrete behaviour in these situations where the MFA (equations (5.3.1) to (5.3.2)) does not provide good agreement. The solid lines correspond to species A, whilst the dashed lines correspond to species B. The averaged discrete results are in black, the MFA in purple and the correlations PDE in cyan. In all of these cases the death rates are zero ($P_d^A = P_d^B = 0$). (a) The two species have the same proliferation rates but different movement rates. We observe that the MFA predicts that both species will behave in exactly the same manner (dashed line for B lies precisely on top of solid line for A). However, the averaged discrete behaviour shows that the species with a higher movement rate has a larger steady-state density. (b) The MFA predicts different behaviour for the two different species whilst the averaged discrete simulations predict almost identical behaviour. (c) The MFA predicts that species B will dominate at steady-state, whilst the averaged discrete simulations predict that species A will dominate.

prediction is not the same as that obtained from the averaged discrete behaviour (Figure 5.3(c)). This shows that even if one species has a lower rate of proliferation, it is still able to maintain a higher long term agent density, on average, if it has a high enough rate of movement compared to the other species. In each of these cases the MFA, which assumes lattice site independence, leads to inaccurate predictions of the averaged discrete results. In these specific cases, the errors can be successfully avoided by incorporating spatial correlations in lattice site occupancy into the model framework, as demonstrated in each of these cases by our correlations PDE model.

Additionally, we look at the behaviour of the correlation functions predicted by our PDE model in Figure 5.4. The auto-correlation functions, F_{AA} and F_{BB}, remain above unity, with similar behaviour to that seen in the previous chapter. This indicates that each species is spatially correlated with itself: if one lattice site is occupied by a

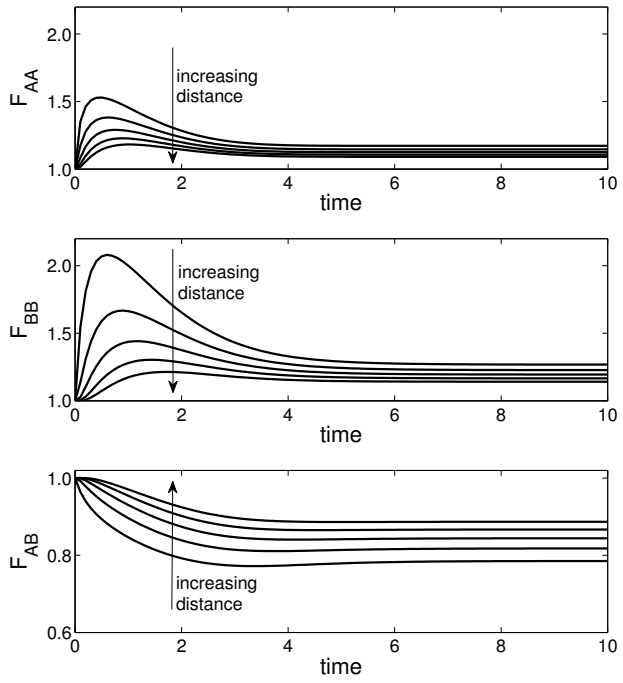


Figure 5.4: The correlation functions F_{AA}, F_{BB} and F_{AB} (from equations (5.3.3) to (5.3.13)), for parameters $P_m^A = 20, P_p^A = 1, P_m^B = 5, P_p^B = 1$ have different properties. The auto-correlation functions (F_{AA} and F_{BB}) are always greater than or equal to unity, indicating that lattice sites are more likely to be occupied by a particular species if there are others nearby. A lower movement rate (as for species B) leads to a species having larger correlations as clusters break up more slowly. The cross-correlation function, F_{AB}, is always less than or equal to unity as the two species are inversely correlated; the presence of A in site l means it is less likely that B will occupy a nearby site. The curves in each plot are for distances from Δ to 3Δ in steps of 0.5Δ.

particular species, it is more likely that the same species will occupy a nearby lattice site. A greater maximum is observed for F_{BB}, for the values of s shown in Figure 5.4, demonstrating that there are stronger correlations between species B than species A. This can be explained by the lower movement rate of species B which leads to slower dispersal of clustered agents. In contrast to what has been seen with all the correlation functions in this thesis thus far, the cross correlation function, F_{AB}, remains below unity, thus the two species are inversely correlated with one another: if one site is occupied by species A, it is less likely that a nearby site will contain species B. As

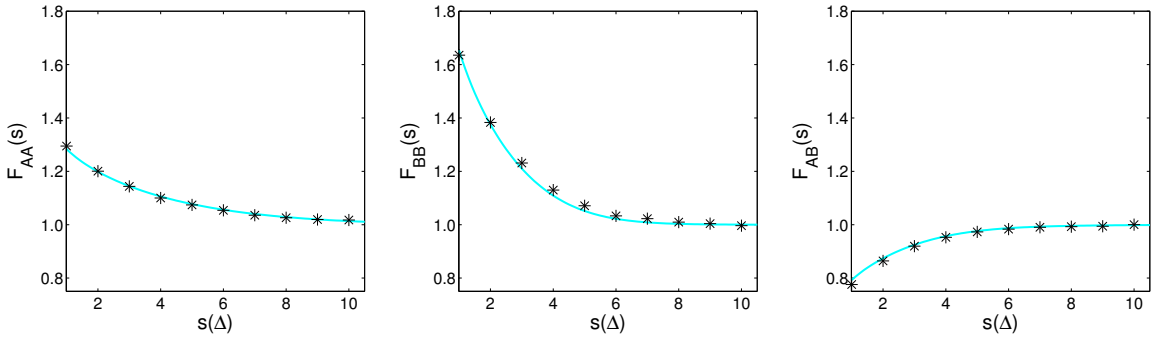


Figure 5.5: We examine the behaviour of the correlations over space at a particular time ($t = 2$) for the case where $P_m^A = 20, P_p^A = 1, P_m^B = 5, P_p^B = 1$ for the correlations PDE method (solid line, using equations (5.3.3) to (5.3.13)). We observe that the auto-correlation functions are above unity for smaller distances, decaying to unity as the distance between the lattice sites increases. The cross-correlation functions are below unity for smaller distances, again approaching unity as the distance between the lattice sites increases. We notice that F_{BB} is larger than F_{AA} as species B has higher levels of correlations due to its lower movement rate. Additionally, we include the averaged discrete results (asterisks), showing that the agreement between the correlations PDE method and the averaged discrete results is excellent.

the distance between lattice sites is increased, all the correlation functions become closer to unity, as correlations decrease with distance.

We also look at how the correlation functions behave over space, comparing the averaged discrete results with our correlations PDE model (Figure 5.5). For the auto-correlation functions, F_{AA} and F_{BB}, we notice that the correlation functions decrease for increasing distances. They eventually tend towards unity, implying that lattice sites far apart are uncorrelated, which is what we anticipate. The correlation function for F_{BB} is generally higher than F_{AA}. This is because B has a lower migration rate, therefore clusters of species B break up more slowly, leading to increased correlations between lattice sites. The cross-correlation function, F_{AB}, also has a value of unity for large distances due to lattice sites becoming effectively independent as the distance between them increases. It is less than unity for smaller distances as site occupancy by the two different species is inversely correlated: if a site is occupied by species A, it is less likely that a nearby site will be occupied by species B. Additionally, we look at

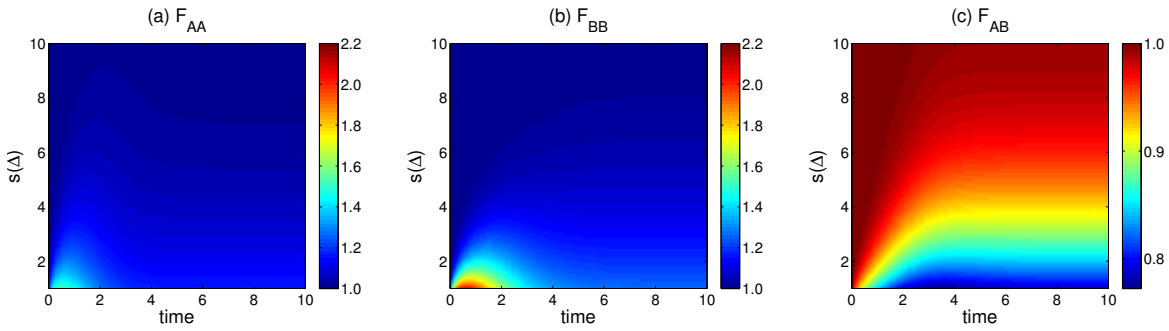


Figure 5.6: The correlation functions vary in both space and time. In this case, we have $P_m^A = 20, P_m^B = 5, P_p^A = P_p^B = 1$. We notice that F_{BB} generally has higher values than F_{AA}, which is due to the lower migration rates of species B: lower migration leads to more clustering, and thus higher levels of correlations. Neither F_{AA} nor F_{BB} go below unity, as lattice sites are correlated when we are taking into account occupancy with the same species. In contrast, F_{AB} never goes above unity as occupancies of lattice sites by two different species is negatively correlated.

the correlation functions from our PDE model over both space and time, confirming the observations we have already made (Figure 5.6).

5.3.5 Including death

Incorporating equal, non-zero death rates for each species into the $N = 2$ system often leads, on average, to the species with the lower steady-state without death now becoming extinct (Figure 5.7), which is the absorbing state for the contact process [Liggett, 1999]. For example, if we include equal non-zero death rates in the system in Figure 5.3(a), then the averaged discrete results show that we expect the species with the lower rate of movement to go extinct. This can be seen in Figure 5.7(a). Additionally, an individual simulation with these parameters can be seen in Figure 5.8, which shows the extinction of species B. Our correlations PDE method agrees with the averaged discrete results, however the MFA predicts that the species will co-exist at the same levels. Moreover, if we include equal non-zero death rates in the system we see in Figure 5.3(c), then the averaged discrete results predict that species B becomes extinct. However, the MFA predicts that species A will go extinct, leaving

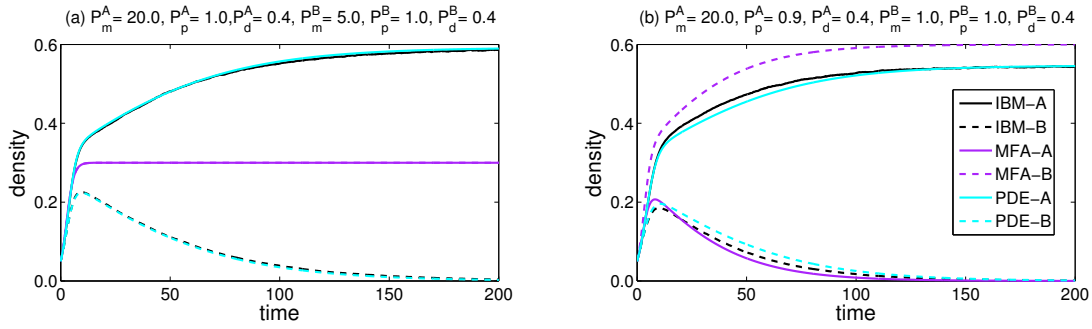


Figure 5.7: Including death in the system can lead to one species becoming extinct for the averaged discrete simulations and the correlations PDE method (using equations (5.3.3) to (5.3.13)). The solid lines correspond to species A, whilst the dashed lines correspond to species B. The averaged discrete results are in black, the MFA in purple and the correlations PDE in cyan. Here we see that the MFA does not agree with the averaged discrete behaviour. (a) The MFA (equations (5.3.1) to (5.3.2)) predicts that both species will coexist with the same steady-state, whilst the averaged discrete model and the correlations PDE method predict that species B will go extinct. (b) The MFA predicts that species A will go extinct whereas the averaged discrete simulations predict that it is in fact species B that will go extinct. The correlations PDE method accurately predicts this behaviour.

species B. Our correlations PDE method accurately predicts that species B will go extinct, providing a good approximation to the averaged discrete results. This can be seen in Figure 5.7(b).

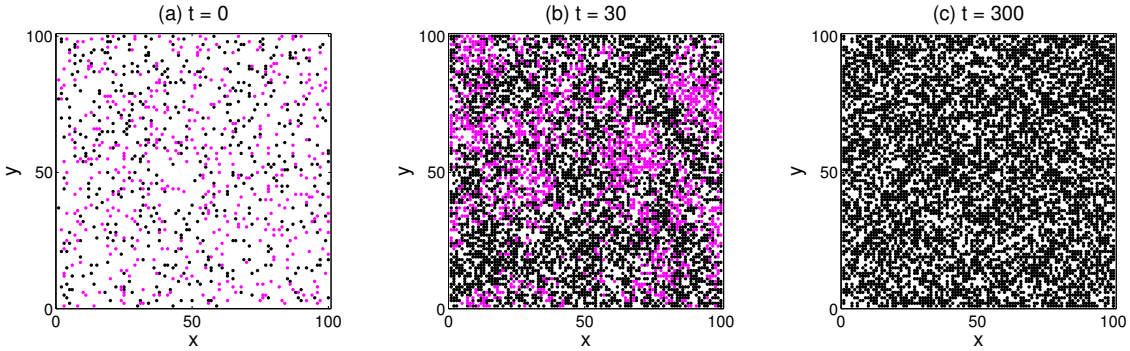


Figure 5.8: A single discrete simulation with non-zero death rates ($P_d^A = P_d^B = 0.4$) shows one of the species becoming extinct. Species A is in black, whilst species B is magenta. In this case, $P_m^A = 20, P_p^A = 1, P_m^B = 5, P_p^B = 1$, corresponding to the averaged behaviour seen in Figure 5.7(a). Whilst the two species have the same proliferation and death rates, the movement rate for species B is lower and leads to species B becoming extinct.

5.3.6 Trading movement for proliferation

We observe in our results thus far that one species may have a lower proliferation rate but still be able to dominate due to a higher movement rate (Figure 5.7). Is it possible for a species to invest more in migration at a proliferative cost to enable it to become the dominant species? We examine the trade-off between proliferation and migration, determing how much of a proliferative cost a species can tolerate for a given set of migration parameters.

We define the switching point as the value of P_p^A, for a given set of other parameters, at which species A first begins to dominate. In Figure 5.9, we vary P_m^A and P_m^B, for $P_p^B = 1$, and observe how the switching point varies. As we would expect, increasing P_m^A relative to P_m^B leads to a lower switching point, as A is more easily able to colonise empty space. Similarly, when P_m^B is high relative to P_m^A, the switching point is greater as species B is now able to move more easily to empty regions and proliferate into them, leaving fewer empty sites available to species A. The switching point is equal to unity when the movement rates are identical (along the diagonal), because the death rates are identical and $P_p^B = 1$, thus we expect the species to behave identically at these points. We can think of the movement as altering the 'effective proliferation rate' of the species, thus although a species may have a lower actual proliferation rate, it has more proliferative opportunities when its migration rate is higher.

5.3.7 Break down of the correlations method

As we have seen, incorporating correlations into a two-species model generally provides excellent agreement with the averaged discrete results. This is especially striking when compared to the predictions made by the MFA, which are sometimes very different from those observed in the averaged discrete case. However, there are times

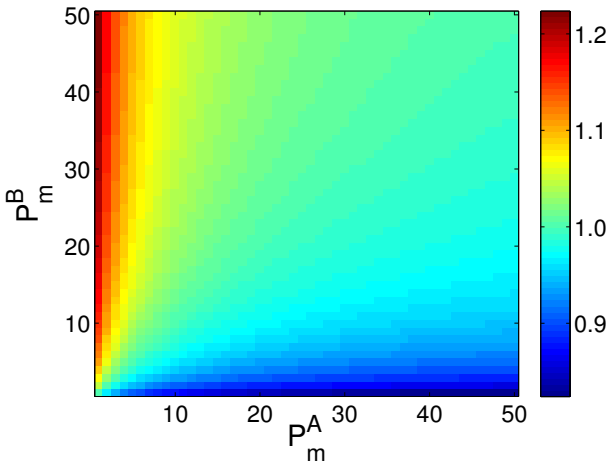
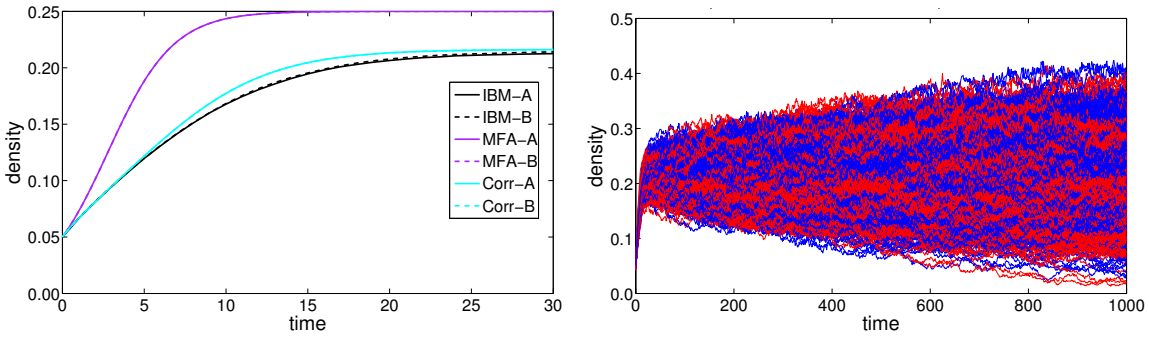


Figure 5.9: The switching point (the value of P_p^A for which species A becomes the dominant species) in our correlations model (using equations (5.3.3) to (5.3.13)) depends on the motility rates of both species, P_m^A and P_m^B. In this case, $P_p^B = 1$ and $P_d^A = P_d^B = 0.1$. As the movement rate of species A increases relative to B, the switching point decreases as A is less clustered thus can proliferate more readily. Similarly, as the movement rate of species B increases relative to A the switching point increases as B is now less tightly clustered, and has more access to empty sites. Along the diagonal, the switching point is equal to unity as all the other parameters for both species are identical at this point.

when the correlations model does not perform as well when approximating the averaged discrete density results. This is to be expected, as we have only gone one step beyond the mean-field and thus, whilst it is an improvement, it is also going to have its shortcomings.

Whilst the correlations model is generally accurate at predicting the averaged discrete behaviour, it does not give us any information about the range of the discrete results. For example, if the parameters for species A and species B are identical, the average behaviour will also be identical. We see this in Figure 5.10(a). However, we find that for a single run in our discrete model, one species will tend to extinction as can be seen in Figure 5.10(b). Each species will go to extinction 50% of the time, thus the average behavior is identical. If we are interested in the range of the discrete behaviour, we would need to incorporate noise in some way, as has been done for some pairwise models [Dangerfield et al., 2009].



(a) Average behaviour (b) Individual simulations

Figure 5.10: The correlations model can accurately predict the averaged discrete behaviour, but does not give any indication of the spread of the individual simulations. In this case, the parameters for species A and species B are identical ($P_m^A = P_m^B = 1, P_p^A = P_p^B = 1, P_d^A = P_d^B = 0.5$). (a) We see that the correlations model provides a significant improvement on the MFA at predicting the averaged discrete behaviour. However, as can be seen in (b), there is a lot of variation in the individual simulations. Here species A is in red and species B is in blue, and 1000 simulations are shown.

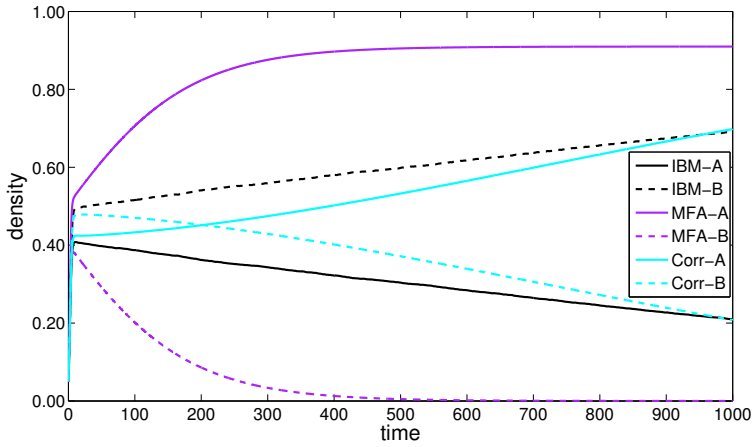


Figure 5.11: In this case the correlations model does not provide a satisfactory approximation to the averaged discrete results. Here the averaged discrete model predicts that species A will go extinct, whilst the MFA (equations (5.3.1) to (5.3.2)) and correlations (using equations (5.3.3) to (5.3.13)) models predict that it will be species B that goes extinct. In this case, $P_m^A = 1, P_p^A = 1.11, P_d^A = 0.1, P_m^B = 5, P_p^B = 1, P_d^B = 0.1$.

For some parameter sets, the correlations model no longer accurately predicts the averaged discrete behaviour (Figure 5.11). We find that this failure is observed close to switching points. We see, in Figure 5.12, that the switching point decreases for increasing values of P_m^A in the averaged discrete case. Thus, if species A is migrating

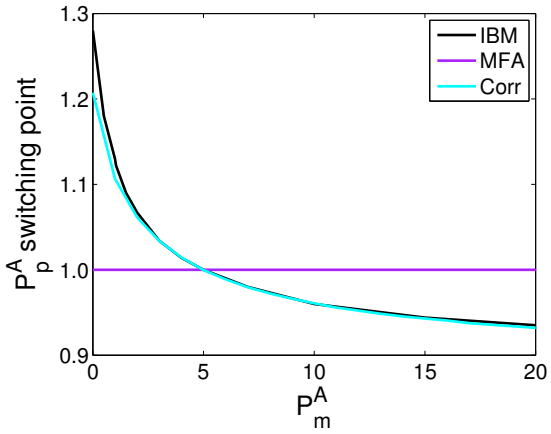


Figure 5.12: The switching point decreases with increasing P_m^A for the averaged discrete case. However, the MFA (equations (5.3.1) to (5.3.2)) predicts the switching point will be independent of P_m^A. The correlations model (using equations (5.3.3) to (5.3.13)) is in good agreement with the averaged discrete behaviour, however we notice some small deviations for low values of P_m^A. In this case, $P_m^B = 5$.

more slowly, it is more likely to go extinct. This behaviour is not reflected in the MFA, for which the switching point is unchanged when P_m^A is varied. The correlations model agrees well with the averaged discrete behaviour, although the curves begin to deviate for low values of P_m^A. Figure 5.11 lies within this region of deviation, and we see that when $P_m^A = 1$, the switching point is slightly lower for the correlations model than for the averaged discrete simulation results. Thus in cases where motility is lower, implying higher levels of clustering, it is more likely that the correlations model begins to make inaccurate predictions.

It is likely that a pair-wise model is not sufficient to capture the underlying behaviour in these regions of parameter space. If this is the case, we would need to include the next level of information by explicitly including the $\rho^{(3)}$ functions and choosing an appropriate closure for any terms involving $\rho^{(4)}$, which would render the model excessively complex.

5.4 Conclusions and future work

In this chapter, we have developed a method for incorporating PDE representations of spatial correlations into a volume-excluding model with multiple species undergoing proliferation, movement, and death on a regular lattice. We showed the success of this method in accurately predicting the averaged discrete behaviour for the two-species case, with and without agent death. Even though the species are not directly interacting, the presence of one species can affect the survival of another. We noted the success of this system in a number of cases where the MFA fails at predicting the averaged discrete behaviour.

The results shown here highlight the dangers of naively using the MFA to model situations where stochastic factors lead to spatial effects being important. We find that the MFA frequently predicts very different behaviour from that seen in our averaged discrete simulations. This could have severe consequences if we use the MFA to make predictions on the behaviour of a system, as the predictions may be completely inaccurate. On the other hand, simply by including PDEs describing lattice site occupancy correlations, we are able to more accurately predict the averaged discrete behaviour in a number of scenarios, leading to a more reliable framework for problems of this kind.

Using our model, we have also examined the trade-off between movement and proliferation that a species may adopt. Whilst one species may have a lower proliferation rate than another, this does not necessarily imply that it will have a lower steady-state density, although this will be the prediction of the MFA. A species can compensate for a lower proliferation rate by having a higher migration rate, thereby altering its effective proliferation as each agent has access to a higher number of unoccupied sites into which it can place daughter agents.

There are some regions of parameter space in which the correlations model does not accurately predict the averaged discrete behaviour. These mainly lie near switch-

ing points where the dominant species changes with a small change in a parameter. In these regions a pair-wise model may no longer be sufficient to capture the underlying behaviour and further extensions may be required to more accurately approximate the averaged discrete results.

Thus far in this thesis, we have observed that the MFA provides a suitable approximation to the averaged discrete behaviour in a small region of parameter space. This can be improved upon by using models which include spatial correlations, with these models performing successfully in a far wider range of parameter space. However, these models also break down under certain conditions. Given experimental data, we would like to be able to determine which method is best suited to a specific biological problem. We have found that the MFA breaks down for slightly clustered systems, and the correlations model for highly clustered systems. Thus measuring the level of clustering in a system could give us an indication of which model is appropriate for given experimental data. In the next chapter, we evaluate and extend some methods from the literature which can be used to determine which model is the most suitable choice for a given situation.

Chapter 6

Clustering measures

The previous chapters have shown that mean-field models often do not accurately predict the averaged discrete density of a system. This can be rectified in many situations by accounting for spatial correlations in lattice site occupancies. However, we have seen that correlations models will also fail under some conditions. Thus, although correlations models are a clear improvement upon mean-field models, they cannot be used to accurately predict the averaged discrete behaviour indiscriminately. Additionally, correlations models are generally more complex to implement than their corresponding mean-field counterparts, thus it would be preferable to use a mean-field model in situations where it provides an accurate approximation of the averaged discrete behaviour. Therefore, we need to be able to determine if the model we are using is the best choice for given experimental data. We know that as the level of agent clustering increases the MFA breaks down, and eventually the correlations approximations do too. Thus, if we have some measure of the clustering of a system we can decide, from the data, which modelling method will be most appropriate. In this chapter, we look at different measures for determining the level of clustering from experimental data for one-species and two-species systems, evaluating their suitability at predicting the best model for a given experimental set-up.

6.1 Introduction

Choosing the most appropriate modelling strategy for a given system leads to accurate results generated in an efficient manner. However, making this choice is not trivial. It would be useful, when presented with experimental data, to have a method to determine which model is most suited to the specific problem in hand. We noted in the previous chapters that, for systems with uniform initial conditions, the development of clusters over time led to failure of the MFA due to the assumption that lattice site occupancies are independent of one another. Using correlations models in these situations leads to more accurate results. However, these only take into account pair-wise densities, relying on an approximation for the triplet behaviour, thus are also liable to failure when clustering becomes extremely high. Thus, measuring the 'patchiness' of a system may indicate the suitability of the mean-field and correlations models [Binder and Landman, 2011; Simpson et al., 2013a, 2014].

In this chapter we use our individual based model to provide the 'data', averaging over 100 realisations, and we discuss how this may differ from real experimental data in Section 6.5. We examine three methods in detail for measuring clustering, considering both the one-species and two-species cases:

- the agent coordination number method (ACN), which considers the number of occupied sites around an agent [Simpson et al., 2014];
- the spatial index method (SI), which looks at the variance in agent number in different regions of the domain [Binder and Landman, 2011; Simpson et al., 2013a];
- the correlations method, which uses correlation functions obtained from the data [Treloar et al., 2014].

We evaluate if our measures can be used to predict the error between the density predictions from our continuum models and the results obtained from our averaged

discrete simulations. We find, in all cases, that we cannot rely on data collected at a single time point to predict the suitability of our models as our measures and prediction errors both change in time, but not at the same rate. We begin, in Section 6.2, with the ACN, which uses the number of occupied sites around an agent to determine how clustered a system is. We then discuss the spatial index method in Section 6.3, before using correlation functions from the data in Section 6.4. Finally, in Section 6.5, we compare the relative success of our methods before concluding in Section 6.6.

6.2 Agent coordination number method

The ACN method considers the occupancies of neighbouring sites around a given agent. This gives an indication of how clustered our agents are, as there will be more occupied neighbouring sites in systems with high levels of clustering. If the system is fairly well-mixed, we should be able to predict how many occupied neighbours an agent will have. In this section, we follow the method outlined by Simpson et al. [2014], whereby we define K as the number of occupied neighbouring sites, using a Moore neighbourhood, around a site that is itself occupied. For a well-mixed system, we expect K to be binomially distributed:

$$Pr(K = k) = \binom{8}{k} c^k (1-c)^{(8-k)}, \quad k = 0, 1, 2, .., 8, \tag{6.2.1}$$

where c is the normalised density. We can compare this predicted distribution to what is actually observed in the results generated by our IBM, quantifying the difference by using a measure of the deviation between the IBM and binomial distribution results:

$$E_k = \sum_{k=0}^{8} |Pr(K = k) - \hat{Pr}(K = k)|, \tag{6.2.2}$$

where $\hat{Pr}(K = k)$ is the discrete distribution, averaged over 100 simulations. To determine how well our continuum approximations are performing, we also need to quantify the difference in the agent density of our continuum models and that obtained from the averaged discrete simulations:

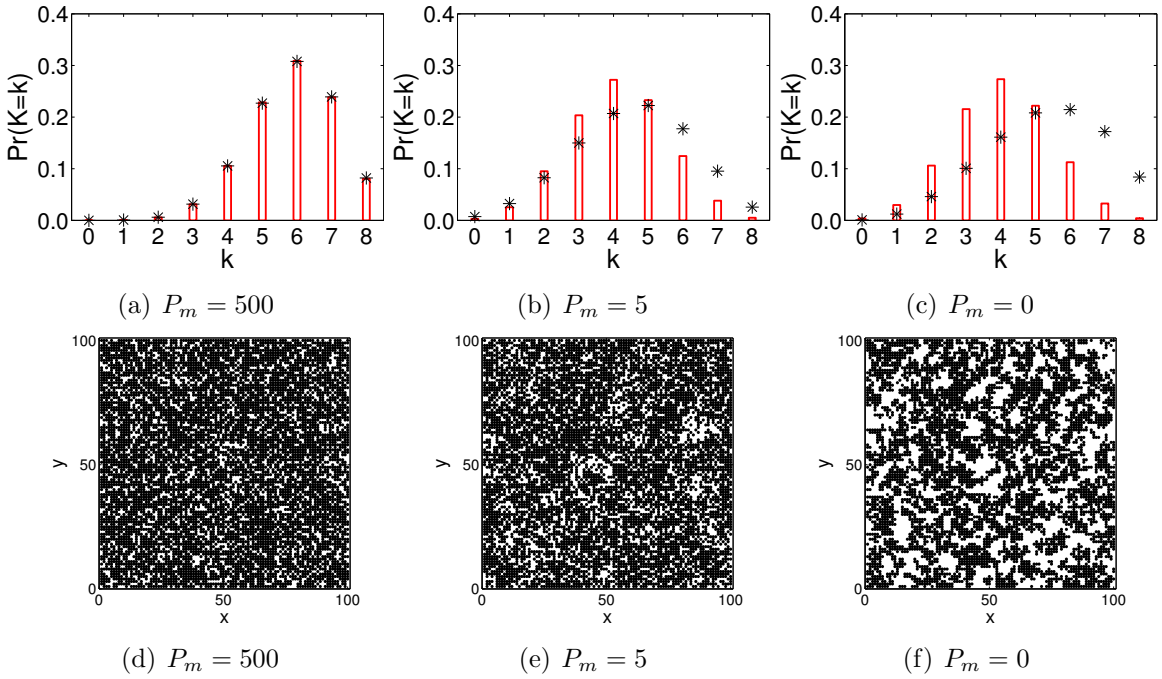
$$E_c = |\hat{c} - c|, \tag{6.2.3}$$

where $\hat{c}$ is the average discrete density and c is the density obtained from one of the continuum models. We use E_{mf} and E_{md} to refer to E_c when using the mean-field and correlations models, respectively, to approximate the density.

6.2.1 One-species system

We begin with the one-species case, whereby we have the same system we discussed in Chapters 2 and 4 (a birth-migration-death process on a regular lattice with random uniform initial conditions). We compare the expected coordination number distribution, equation (6.2.1), with the same measure obtained from one hundred discrete simulations. We begin with a situation where we expect them to be in agreement, i.e. in a well-mixed system. The agreement in this case is excellent, Figure 6.1(a). As we lower the movement rate, we observe that the agreement between the actual and predicted results worsens, Figures 6.1(b) and 6.1(c). Lower levels of movement imply higher levels of clusters which result in the distribution being skewed towards higher values of coordination numbers for the averaged discrete results. However, this is not observed in the theoretical distribution, which assumes the system is well-mixed.

In Simpson et al. [2014], the authors choose to look at the data at a specific point in time when evaluating the level of clustering. As we expect the degree of clustering to vary over time, we instead examine the time-series behaviour of E_k. We first consider the behaviour when $P_d = 0$, in Figure 6.2(a). We see that E_k begins

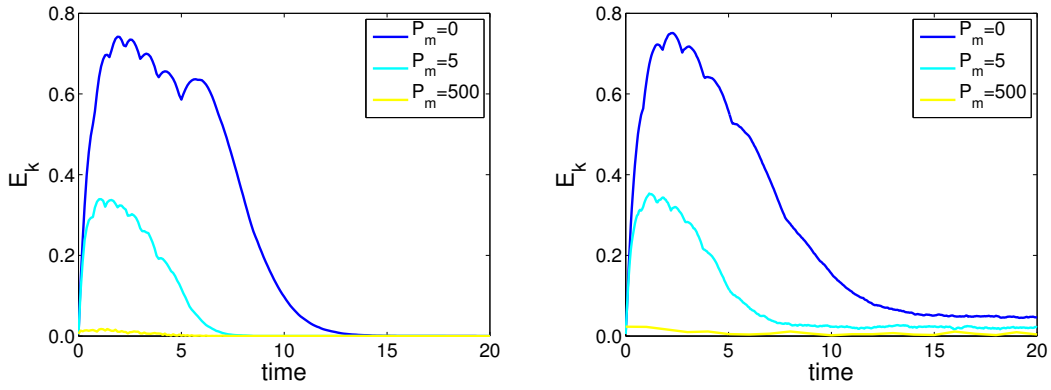


(a) $P_m = 500$ (b) $P_m = 5$ (c) $P_m = 0$

(d) $P_m = 500$ (e) $P_m = 5$ (f) $P_m = 0$

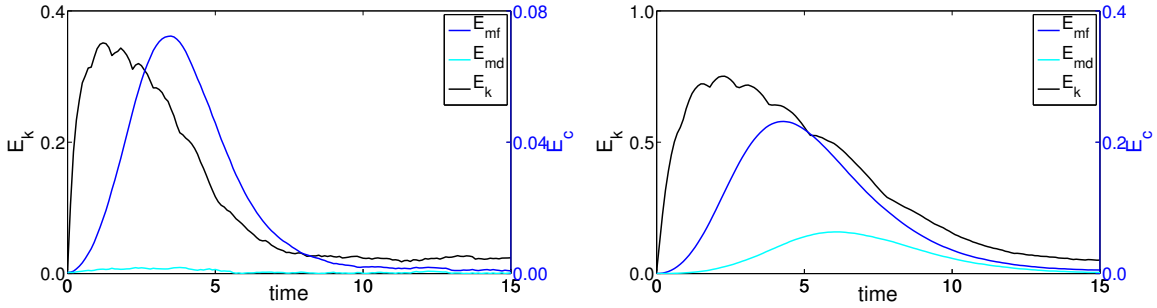
Figure 6.1: The agreement of the ACN calculated from the discrete data (black asterisks) and equation (6.2.1) (red bars) decreases with the movement rate. (a) High values of P_m give a well-mixed system and, therefore, we obtain excellent agreement between the theoretical distribution and the averaged discrete results. (b)-(c) As the movement rate is lowered we see that our theoretical distribution and averaged discrete results do not display good agreement. In (d)-(f), we see snapshots of a discrete simulation corresponding to (a)-(c), respectively. These demonstrate the clustering for each set of parameters. For all cases, $P_i = 0.1, P_p = 1, P_d = 0.1$ and we are looking at the distribution at $t = 4$. The number of agents on the lattice at this time depends on the parameters used: in (a) and (d), the lattice is 76% occupied; in (b) and (e), the lattice is 71% occupied; in (c) and (f), the lattice is 55% occupied.

at zero, as the distribution is initially random. It then increases to a maximum, before decreasing back to zero. The steady-state value of E_k is zero as, when there is no death in the system, the lattice eventually becomes completely occupied. As expected, the maximum value of E_k (E_{kmax}) is higher when the rate of movement is lower due to the greater degree of clustering. In the case where $P_d \neq 0$, Figure 6.2(b), the behaviour is similar but the steady-state value is not necessarily zero as the lattice does not become completely occupied. Due to the large variation of E_k over time, we need to be careful using E_k as an indicator of the system's clustering if we only consider one time-point, as it may not be indicative of the overall behaviour.



(a) $P_d = 0$ (b) $P_d = 0.1$

Figure 6.2: The difference between the observed and theoretical predictions of the agent coordination number, E_k, equation (6.2.2), depends on time. For lower migration rates, the deviation is larger, as more clusters are observed. (a) When there is no death, the final behaviour matches the mean-field as the lattice is completely filled. (b) When $P_d \neq 0$, the steady-state for E_k is greater than 0 as the lattice will not be completely filled. The parameters in all cases are $P_p = 1, P_i = 0.1$.



(a) $P_m = 5$ (b) $P_m = 0$

Figure 6.3: There is a delay between the time taken to reach the maximum of E_k, equation (6.2.2), and that taken to reach the maximum of the deviation from the density data, E_c, equation (6.2.3), where E_{mf} and E_{md} refer to approximating the continuum density in E_c with the mean-field and correlations models, respectively. The axis on the left corresponds to E_k, whilst that on the right corresponds to E_{mf} and E_{md}. The error in predicting the averaged discrete density for the mean-field, E_{mf}, and for the correlations model, E_{md}, also do not have coinciding maxima. In both figures, $P_p = 1, P_d = 0.1, P_i = 0.1$.

We now examine the difference between the continuum density predictions and the averaged discrete density results, E_c, using equation (6.2.3) and determine how this measure relates in time to E_k. In Figure 6.3, we observe that there is a delay between the maximum of E_k and the maxima of E_c for both the MFA and correlations

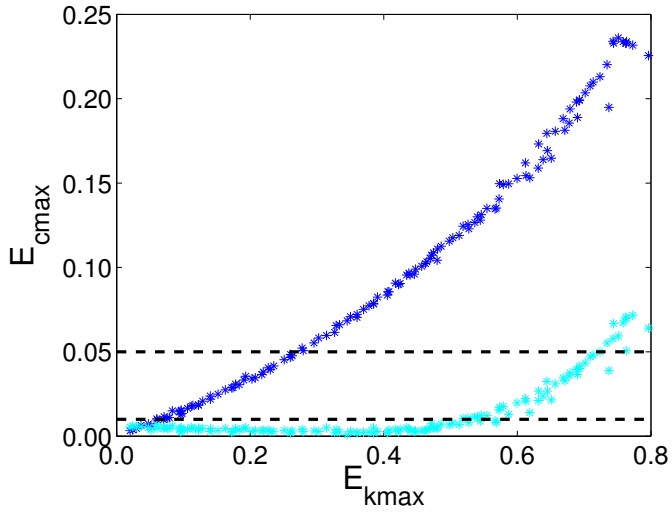
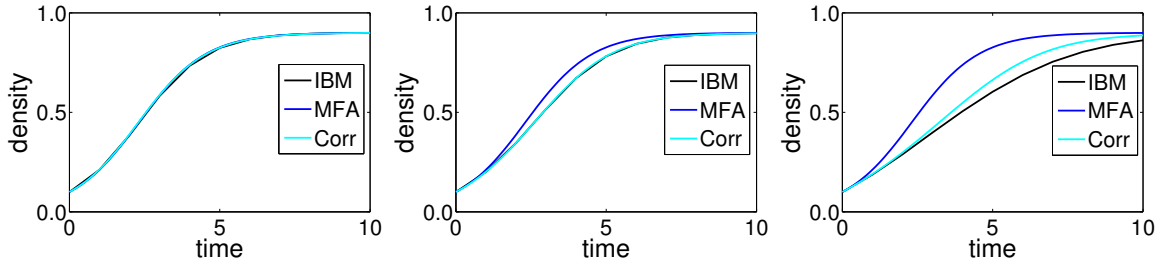


Figure 6.4: There is a clear trend between the maximum values of E_k, equation (6.2.2), and E_c, equation (6.2.3), for both the mean-field (blue) and correlations (cyan) models. Each point represents the values of E_{cmax} and E_{kmax} for a specific set of parameters: we vary P_m from 0 to 1000, P_p from 0 to 10 and P_d from 0 to P_p. These results indicate that it is feasible to use E_{kmax} to indicate which model might best suit given experimental data, as we can estimate the error in each continuum model if we know E_{kmax}. The deviation between the averaged discrete and continuum approximations increases as E_k increases. The maximum error for the mean-field model increases much more rapidly than it does for the correlations model, as we expect. The correlations model provides more accurate predictions of the averaged discrete density for higher levels of clustering, whereas the mean-field model breaks down rapidly as the system becomes patchy. The dashed lines indicate the 1% and 5% thresholds in the maximum density estimation error.

model. Thus, we must be cautious about using clustering measures on experimental data at only single time points to predict the error in our models as results could be misleading. Rather, we should consider the clustering measures and model errors over time and examine the maxima to determine if there are consistent relationships between these values. We see, in Figure 6.4, that there are clear trends between the maxima of E_k and E_c for both the mean-field and correlations models. From this we can determine threshold values of E_{kmax} which can be used to predict the most appropriate model for a given circumstance. For example, using Figure 6.4, if our E_{kmax} value is 0.2, we would use the correlations model if we want an error of less



(a) $P_m = 500$ (b) $P_m = 5$ (c) $P_m = 0$

Figure 6.5: Density plots corresponding to different regions of Figure 6.4. (a) $E_k = 0.024$, thus E_{mf} and E_{md} are both below 1%; both the mean-field (equation (2.4.3)) and correlations models (using the equations in Section 4.3) provide a suitable approximation to the averaged discrete behaviour. (b) $E_k = 0.35$, thus $E_{mf} > 5\%$ and $E_{md} < 1\%$; the correlations model provides a suitable approximation to the averaged discrete behaviour but the mean-field does not. (c) $E_k = 0.75$, thus both E_{mf} and E_{md} are over 5%; neither continuum model provides a suitable approximation to the averaged discrete results.

than 1% but our MFA if we only require an error below 5%.

In Figure 6.5, we look at some specific cases whereby we determine which model is best suited to a set of parameters as determined from the ACN applied to the averaged discrete data. We then confirm this by examining the averaged discrete, MFA and correlations density results, finding that the errors agree with what we expect.

Using the ACN method on a one-species system can provide an indication of the best model to use for a given situation. Of course, using this method relies upon us knowing the maximum value of E_k, which we may not always obtain depending on how often our experimental data are recorded. Thus the ACN as a clustering measure may not be a robust method if we cannot take sufficiently frequent experimental readings. Our experimental data may also have more than one species present, and we now investigate if this method can be successfully extended to the two-species case.

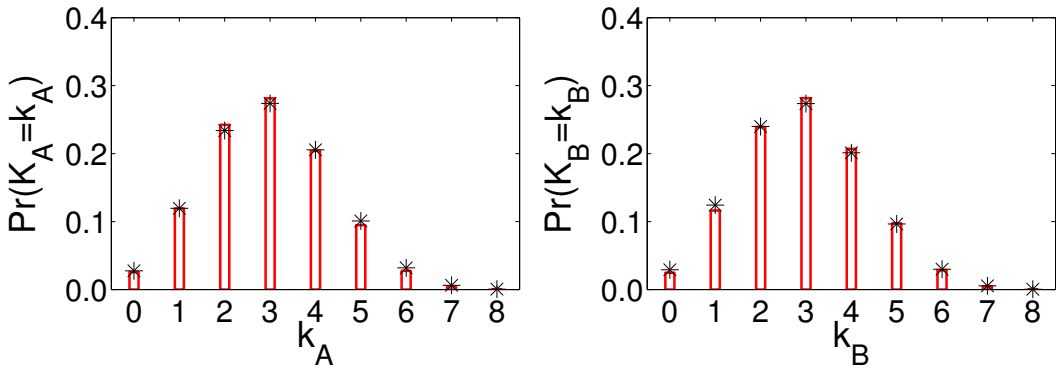


Figure 6.6: In the well-mixed case, we observe good agreement between the averaged discrete results (black asterisks) and the theoretical prediction, using equation (6.2.4) (red bars), in the two-species case. In this case, $P_m^A = P_m^B = 500, P_p^A = P_p^B = 1, P_d^A = P_d^B = 0.1, P_i^A = P_i^B = 0.05, t = 5$.

6.2.2 Two-species system

The ACN provides a fairly good indication as to which model is best suited to a one-species system, if we have sufficient data over time. We now turn to the case with two-species, the same system described in Chapter 5. We modify the ACN method described in the previous section by now considering the ACN for the I^{th} species (the number of agents of the same species around a chosen agent):

$$Pr(K_I = k_I) = \binom{8}{k_I} c_I^{k_I}(1 - c_I)^{(8-k_I)}, \quad k_I = 0, 1, 2, ..., 8. \tag{6.2.4}$$

We can examine this theoretical prediction against the averaged discrete ACN for each species. We again begin with the well-mixed case, confirming that the expected and actual coordination number match up well in these situations (Figure 6.6). If we lower the movement rate for one of the species, the predicted coordination number for both species begins to differ from that predicted in the well-mixed case. This means that even though it may be appropriate to use the MFA for a species when it is alone, it may no longer be suitable once another species is added (Figure 6.7).

When considering experimental data, it may be easier if we do not need to distinguish between the different species types. We therefore use equation (6.2.1) where

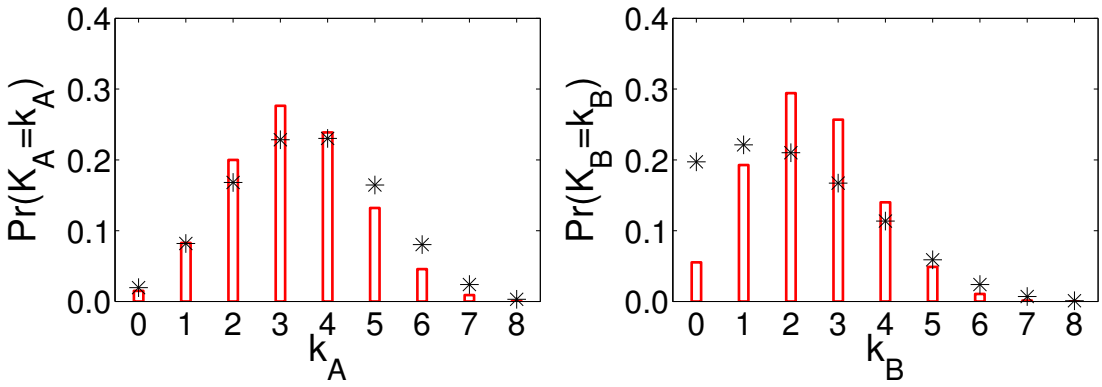
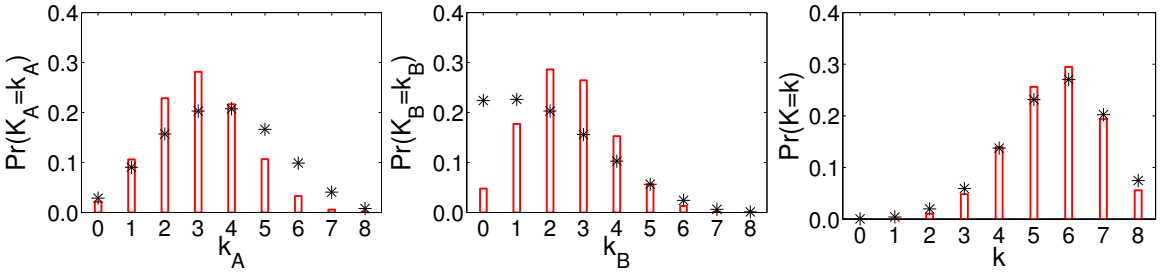


Figure 6.7: Although one species alone has a good match with the actual and theoretical agent coordination numbers, adding another species to the system may alter this agreement. Species A has parameters identical to those used in Figure 6.1(a) where the theoretical and actual agent coordination numbers are in good agreement. In this case we have mixed species A with another species with a lower rate of migration: $P_m^A = 500, P_m^B = 5, P_p^A = P_p^B = 1, P_d^A = P_d^B = 0.1, P_i^A = P_i^B = 0.05, t = 5$. Species A is more clustered according to the simulation results than predicted by the theoretical distribution. This is because species B acts as a 'blockage' to species A, thus it cannot disperse as easily. Species B is less clustered according to the simulation results than the theoretical distribution predicts because species B is restricted by the presence of species A thus has a limited number of unoccupied neighbours into which it can proliferate.



(a) (b) (c)

Figure 6.8: We need to be able to distinguish between our species for the ACN to be meaningful. (a), (b) We see that for K_A and K_B, the ACN predicted theoretically (red bars) using equation (6.2.4) does not match that obtained from the averaged discrete simulations (black asterisks). (c) If the species are indistinguishable, $c = c_A + c_B$, then the theoretical and discrete predictions for the ACN are in reasonable agreement. In this case, $t = 5$, $P_m^A = 20, P_m^B = 5, P_p^A = P_p^B = 1, P_d^A = P_d^B = 0.1$.

we allow our species to be indistinguishable, thus $c = c_A + c_B$. We see, in Figure 6.8, that whilst the predicted number of neighbours using the averaged discrete data and well-mixed theoretical distribution, equation (6.2.1) make different predictions

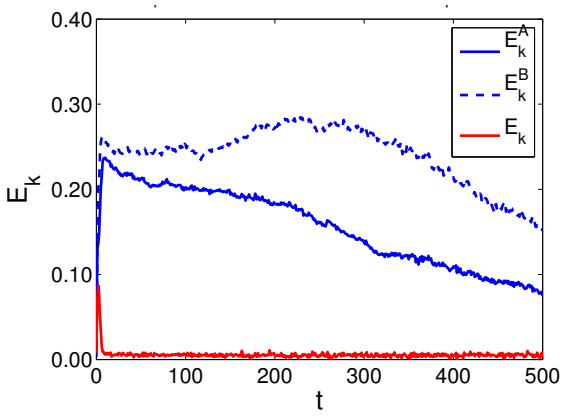


Figure 6.9: The deviation from the theoretical distribution, E_k^A and E_k^B using equation (6.2.5), differs between the two species if they have different parameter values. If the species are indistinguishable experimentally, E_k is very low except for small times. In this case, $P_m^A = 20, P_m^B = 5, P_p^A = P_p^B = 1, P_d^A = P_d^B = 0.1$.

for each species alone (using equation (6.2.4)), they make the same prediction when the species are indistinguishable. Thus it is essential that we are able to distinguish between our species if there is more than one species present when we are using the ACN, otherwise our results may be meaningless.

We again use E_k to provide a measure of the deviation between the observed and theoretical curves, but we now need to do so for each species, E_k^I, with the equation being:

$$E_k^I = \sum_{k=0}^{8} |Pr(K^I = k^I) - \hat{Pr}(K^I = k^I)|, \tag{6.2.5}$$

If the species are indistinguishable, we have only one E_k, in the same way as the one-species case. We observe how E_k^I behaves over time in Figure 6.9. We see that the species with the higher movement rate will have a lower E_k. This is to be expected as a faster moving species will generally be less clustered. We also note that when the species are indistinguishable, E_k is very low except for early times, again implying that we must be able to distinguish between our cell types experimentally for us to use the ACN to determine which model is most suited to a set of data. Looking at the errors in the density predictions in Figure 6.10, we see that these are very similar

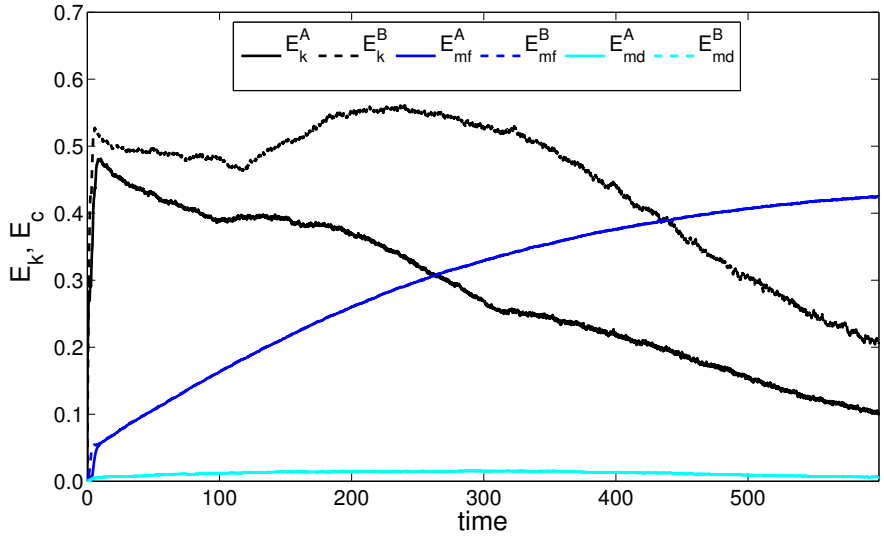


Figure 6.10: The deviation of the continuum predictions (MFA and correlations model) from the averaged discrete results (E_{mf} and E_{md}, respectively), using equation (6.2.3), are very similar for the two different species. In this case the dashed and solid lines are virtually indistinguishable. However the errors in the ACN (E_k^A and E_k^B, from equation (6.2.5)) are not similar. Thus we cannot indiscriminately choose one of the two species to evaluate as doing so would lead to different threshold results when comparing the level of clustering to the error at predicting the averaged discrete behaviour. In this case, $P_m^A = 20, P_m^B = 5, P_p^A = P_p^B = 1, P_d^A = P_d^B = 0.1$.

for the two species, even though E_k^A and E_k^B are different. This means we cannot just follow one species as it will not be able to provide a clear trend between E_k^I and E_c^I. We choose to take the highest maximum value from the two species as our measure of E_{kmax}, thus $E_{kmax} = \ textrmmax(E_{kmax}^A, E_{kmax}^B)$ for two species, A and B. Similarly $E_{cmax} = \max(E_{cmax}^A, E_{cmax}^B)$.

We now determine if there is a relationship between E_{kmax} and E_{cmax} similar to that seen in the one-species case. We focus on the case where $P_d = 0$ as evaluating our models to steady-state in the discrete simulations when $P_d \neq 0$ is time-consuming. Even when distinguishing between species, there does not seem to be a clear trend such as that observed in the one-species case. This can be seen in Figure 6.11, where we see a large spread in the data points.

Not only can we not easily determine which continuum model is best when we

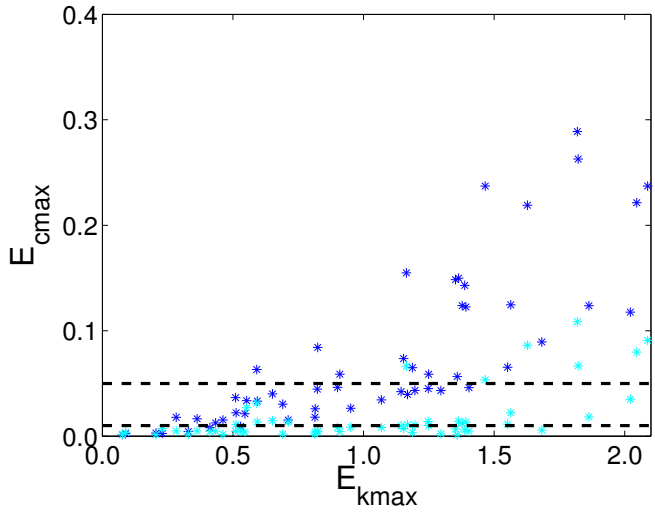


Figure 6.11: There is not a clearly defined relationship between E_{cmax} for the MFA (blue) and correlations model (cyan) and E_{kmax} for the two-species case, although there was a clear trend for the one-species case in Figure 6.4. Thus the ACN is not a robust method for determining a suitable model for two-species data. The dashed lines indicate the 1% and 5% error thresholds.

are faced with a two-species system, we are also not guaranteed that the models that can successfully be used in the one-species case will be best when species are mixed together. Determining the best model for the one-species case does not necessarily imply this model will be the same when two species are combined. The MFA might be appropriate for each species alone, but not be appropriate when the species are mixed. For example, for a cutoff of 1%, we find that species A ($P_m = 100, P_p = 1, P_d = 0$) and species B ($P_m = 100, P_p = 0.2, P_d = 0$) both give $E_{mf} < 1\%$. However, if we mix species A and B together, $E_{mf} > 1\%$. Thus, even knowing that two species alone can successfully be modelled by a specific continuum model does not guarantee that a mixture of the species can be modelled using the same strategy.

6.3 Spatial index method

A different method which can provide information about the patchiness of a system uses a scaled spatial variance [Binder and Landman, 2011] as has been applied to

experimental images [Simpson et al., 2013a]. This method examines the variance in the number of agents in different regions of the domain, thus indicating whether the same number of agents on average occupy a given area. In this section, we follow the method presented by Binder and Landman [2011] to determine the spatial index. We begin by dividing our 2D domain, consisting of $N_x \times N_y$ sites, into M equal-sized bins. We expect the average bin count to be $Q(t)/M$ where $Q(t)$ is the total number of agents on the lattice at time t ($Q(t) = cN_xN_y$). We can calculate the variance of the bin count as

$$\text{var} = \frac{1}{M}\sum_{j=1}^{M}\left(b(j) - \frac{Q}{M}\right)^2, \tag{6.3.1}$$

where $b(j)$ gives the number of cells in bin j. There are two extreme cases: (i) each bin contains exactly the average number of agents (var = 0), and (ii) all agents reside in one bin. In the latter case, we can write

$$b(j) = \begin{cases} Q, & j = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{6.3.2}$$

Therefore equation (6.3.1) in this case becomes

$$\text{var}_0 = Q^2\left(\frac{M-1}{M^2}\right). \tag{6.3.3}$$

We use this to scale our variance, giving a spatial index

$$I = \frac{\text{var}}{\text{var}_0}. \tag{6.3.4}$$

Thus, the spatial variance index is defined as:

$$I(t) = \frac{M}{Q(t)^2(M-1)}\sum_{j=1}^{M}\left[b(j,t) - \frac{Q(t)}{M}\right]^2. \tag{6.3.5}$$

Randomly distributed agents will not be completely evenly distributed, thus we need to approximate the deviation we expect if the system is completely random. We do so using the complete spatial randomness state, $I_{csr}(t)$ [Binder and Landman, 2011], which, when the size of the cells (s) is much smaller than the size of the domain, can be approximated as:

$$I_{csr}(t) \approx \frac{1}{Q(t)} - \frac{s}{N_x N_y}, \tag{6.3.6}$$

as defined by Binder and Landman [2011], where $s = 1$ in our case. Comparing $I(t)$ and $I_{csr}(t)$ can be used to provide an indication of the patchiness of the system: when $I(t) \leq I_{csr}(t)$, the system is close to random, whereas when $I(t) > I_{csr}(t)$ the system is more clustered than we would expect from a completely random system [Simpson et al., 2013a].

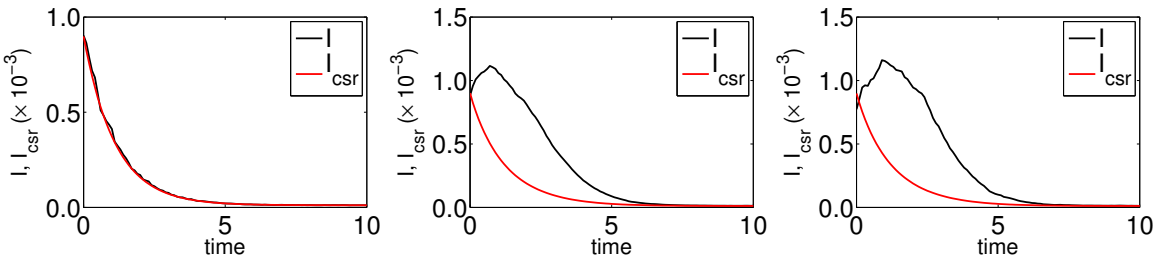
6.3.1 One-species system

We begin with the one species case, examining the spatial index for a range of parameters. We compare the spatial index to the complete spatial randomness (CSR) state for different values of P_m in Figure 6.12. We note that higher rates of movement lead to the spatial index being closer to the CSR state.

Naturally, I may vary depending on the number of bins into which we divide our domain. We vary our number of bins, M, to determine the effect this has upon I. This can be seen in Figure 6.13. We find that as M increases we approach the CSR state. When $M = N_x N_y$, I is identical to the CSR. By substituting $M = N_x N_y$ into equation (6.3.1), we obtain

$$\text{var} = \frac{1}{N_x N_y} \sum_{j=1}^{N_x N_y} \left(b_j - \frac{Q}{N_x N_y} \right)^2. \tag{6.3.7}$$

Each bin will have either one or zero agents in it, thus we can rewrite equation (6.3.7)



(a) $P_m = 5000$ (b) $P_m = 50$ (c) $P_m = 5$

Figure 6.12: As P_m decreases, (a)-(c), the spatial index, given by equation (6.3.5), deviates more from the CSR state, equation (6.3.6). (a) When movement is very high, the variance in bin occupancy is very similar to what we would expect if the agents were randomly distributed. (b)-(c) Decreasing the rate of migration leads to larger discrepancies between I and I_{CSR} as agents tend to be more clustered than we would expect in the random case thus leading to a larger variance in bin occupancy. In all cases, $P_p = 1, P_d = 0.1$.

as

$$\text{var} = \frac{1}{N_x N_y}\left(\sum_{j=1}^{Q}\left(1 - \frac{Q}{N_x N_y}\right)^2 + \sum_{j=1}^{N_x N_y - Q}\left(\frac{Q}{N_x N_y}\right)^2\right), \tag{6.3.8}$$

which simplifies to

$$\text{var} = \frac{Q}{N_x N_y}\left(1 - \frac{Q}{N_x N_y}\right). \tag{6.3.9}$$

Performing the same scaling as before, and assuming $N_x N_y \gg 1$, we obtain the following expression for the spatial index:

$$I = \frac{1}{Q} - \frac{1}{N_x N_y}, \tag{6.3.10}$$

which is identical to the CSR state, confirming what we have observed in Figure 6.13. The sensitivity of I to M is problematic when applying this measure. For the rest of the chapter, we use $M = 25$ as our bin number.

We next examine how the error in our continuum predictions, E_{mf} and E_{md}, and the spatial index vary in time. We define our spatial index measure as the difference between I and I_{CSR},

$$SI_{max} = \max(I - I_{csr}). \tag{6.3.11}$$

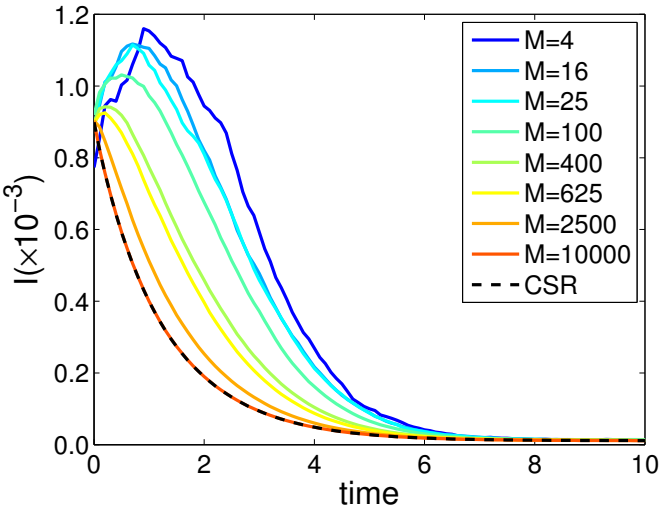


Figure 6.13: As the number of bins, M, increases, the spatial index, given by equation (6.3.5), approaches the CSR state, given by equation (6.3.6). When M is equal to the number of lattice sites, I is identical to the CSR - the black dashed line coincides precisely with the $M = 10,000$ line.

We observe, in Figure 6.14, that the maximum of SI occurs before the maxima of the errors in our continuum approximations. This is the same as what we observed in the ACN method from the previous section. Thus we must again be cautious about using this method to distinguish between models using data collected at only one snapshot in time as it may not be truly indicative of the behaviour at all times. We again examine if there is a clear trend between the maximum values of SI and the errors to determine whether either of the continuum methods is suitable.

We see in Figure 6.15 that this measure does not give a clear relationship between E_{cmax} and SI_{max} as SI_{max} becomes quite large. The behaviour is not so noisy for lower SI_{max} thus we can probably determine a cutoff for the 1% error case for the MFA ($SI_{max} < 5.8 \times 10^{-4}$).

6.3.2 Two-species system

Although the SI method did not perform well in the one-species case, we extend it to account for two species, to confirm it does not adequately predict which model to

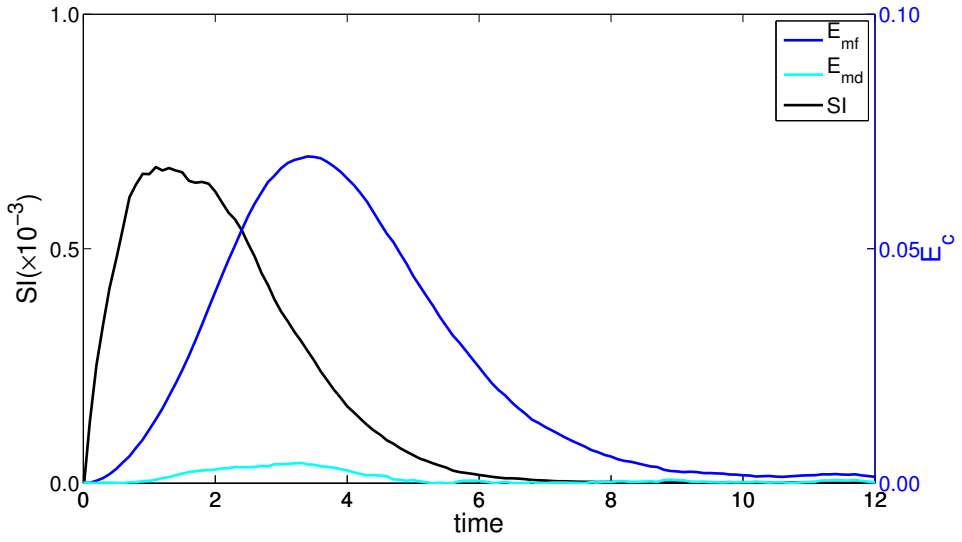


Figure 6.14: There is a delay between the maximum of the spatial index, given by equation (6.3.4), and the maximum of the errors, given by equation (6.2.3), in estimating the averaged discrete behaviour by the continuum models (E_{mf} for the mean-field and E_{md} for the correlations model). In this case $P_m = 5, P_p = 1, P_d = 0.1$.

choose in this case. We now count the number of each cell type in a given area of the domain:

$$I^A(t) = \frac{M}{Q(t)^2(M-1)} \sum_{j=1}^{M} \left[b(j,t) - \frac{Q^A(t)}{M} \right]^2, \tag{6.3.12}$$

and

$$I^B(t) = \frac{M}{Q(t)^2(M-1)} \sum_{j=1}^{M} \left[b(j,t) - \frac{Q^B(t)}{M} \right]^2, \tag{6.3.13}$$

where Q^I is the number of agents of species I. We alter the complete spatial randomness state in the same way:

$$I^A_{csr}(t) \approx \frac{1}{Q^A(t)} - \frac{s}{N_x N_y}. \tag{6.3.14}$$

and

$$I^B_{csr}(t) \approx \frac{1}{Q^B(t)} - \frac{s}{N_x N_y}. \tag{6.3.15}$$

As with the one-species case, we compare the spatial indices to what we would expect

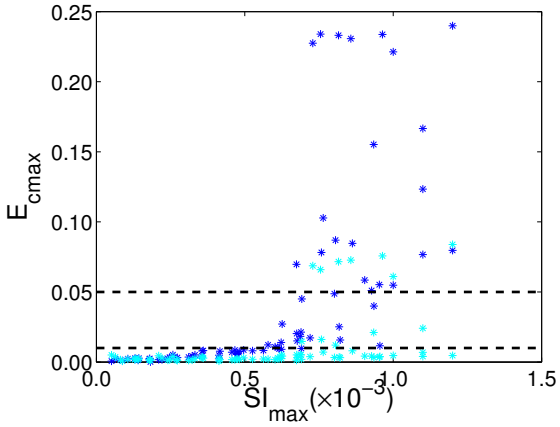


Figure 6.15: There is no significant trend between the maximum error of the continuum approximations (MFA in blue and correlations in cyan) and SI_{max} when SI_{max} is large, where SI_{max} is the maximum of the difference between I and I_{csr}. Dashed lines correspond to 1% and 5% errors. We see that when SI_{max} is small, the relationship between the maxima is not so noisy thus we can determine the 1% error threshold for the MFA as being $SI_{max} = 5.8 \times 10^{-4}$.

in the CSR state. We once more begin with the well-mixed case, Figure 6.16, where we observe that the spatial indices are in good agreement with the CSR. If we lower the movement rates, Figure 6.17, we find that the indices begin to deviate from what is expected in the random state. Even if one species would be close to the random state alone, adding another species alters this as it acts as a hindrance, Figure 6.18.

We define $SI^A_{max} = \max(I^A - I^A_{csr})$ and $SI^B_{max} = \max(I^B - I^B_{csr})$. We define our overall measure, SI_{max}, as the maximum of SI^A_{max} and SI^B_{max}. We again determine if there is a relationship between SI and the maximum error in our continuum model approximations of the averaged discrete behaviour. We see, in Figure 6.19, that we again do not have a particularly clear relationship between SI_{max} and the maximum errors in our continuum approximations of the averaged discrete behaviour. We can determine the cutoff for $< 1\%$ error in our MF prediction: this is seen to be 1.3×10^{-3}. We cannot with any sureness predict any other threshold values.

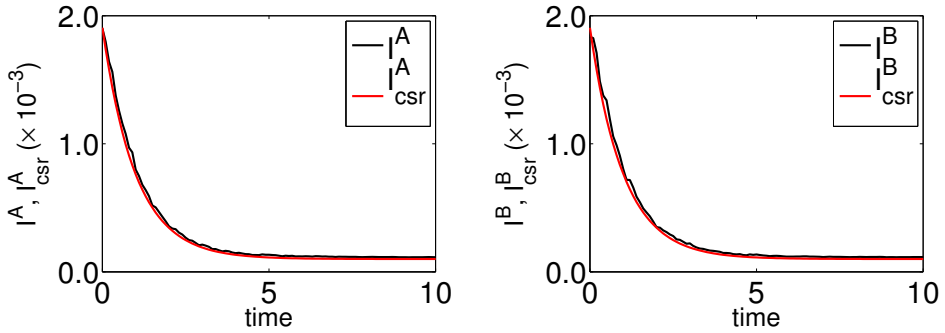


Figure 6.16: When migration rates are high our spatial indices in the two-species case, equations (6.3.12) and (6.3.13), are again close to the CSR states, equations (6.3.14) and (6.3.15). In this case, $P_m^A = P_m^B = 5000, P_p^A = P_p^B = 1, P_d^A = P_d^B = 0$.

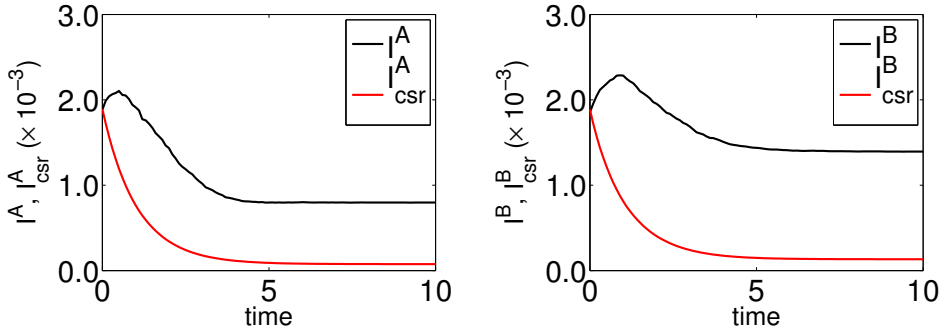


Figure 6.17: In the two-species case, the spatial variance index for the species with the higher movement level lies closer to the CSR as this species will be less clustered. In this case, $P_m^A = 50, P_m^B = 5, P_p^A = P_p^B = 1, P_d^A = P_d^B = 0$.

6.4 Spatial correlations method

The correlations in lattice site occupancy can indicate how clustered a system is. In previous chapters, we compared the correlation functions from our PDE with those determined directly from the data. We now consider if these same correlation functions estimated from the data give any indication of which model is most suited to a particular experimental data set. We calculate the correlation functions from the data by first calculating the $\rho^{(2)}$ functions at the distance in question. For example, to obtain $F_{AA}(\Delta)$, we find $\rho^{(2)}(A_l, A_{l+\Delta})$ on our lattice, by determining how many neighbouring pairs of sites both contain agents of species A. We then divide by c_A^2 to

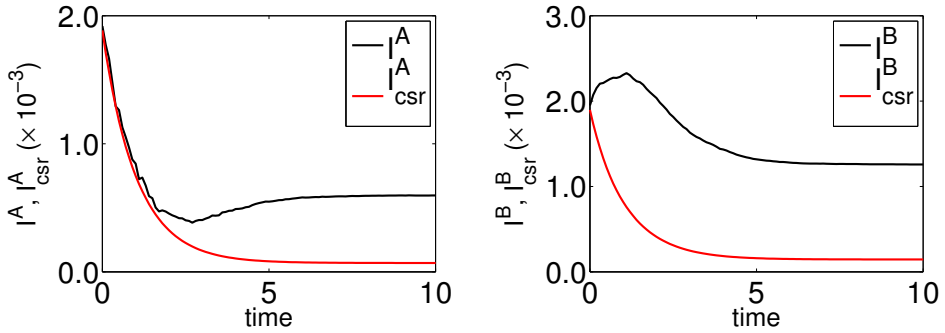


Figure 6.18: Although the spatial index may be similar to the CSR for a species when it is alone, this does not necessarily hold if two species are mixed together. In this case, species A is the same as that used in Figure 6.12(a), where I was seen to be similar to I_{csr}. Species B, which has a lower rate of migration, has now been added to the system and we observe that I^A is now further from the CSR. In this case, $P^A_m = 5000, P^B_m = 5, P^A_p = P^B_p = 1, P^A_d = P^B_d = 0$.

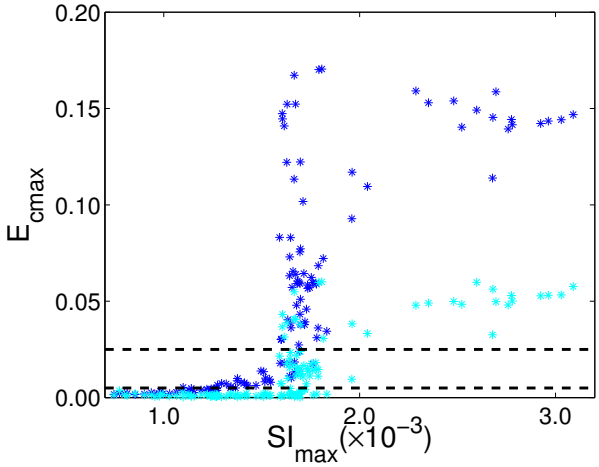


Figure 6.19: Comparing SI_{max} with the maximum errors in our continuum predictions (MFA in blue, correlations in cyan) when predicting the averaged discrete simulation results does not give a significant trend for large SI_{max}. We can use this measure to determine when the MFA error is below 1%, but not for any other threshold values as there is too wide a spread in the points.

obtain the relevant correlation function. In this section we look at this method, again considering both the one-species and two-species cases, examining the correlation functions at distances of $\Delta, 2\Delta$ and 3Δ.

6.4.1 One-species system

For the case with a single species, we consider three correlation functions: $F_{AA}(\Delta)$, $F_{AA}(2\Delta)$ and $F_{AA}(3\Delta)$. We observed in previous chapters that the correlation functions vary with time. As with the previous methods, we want to know how this variation relates to the time-series behaviour of the error between density predictions of the continuum models and the averaged discrete results. In Figure 6.20, we again observe a delay between the maxima of the correlation functions and of the error in our continuum predictions. Thus, we once more cannot rely solely on a single point in time but must consider the time-series behaviour. As with the previous two methods we discussed, we examine if there is a relationship between the maxima of the correlation functions and those of the errors in the continuum predictions. In Figure 6.21 we examine if there is a trend between the maximum of the correlation functions (at $\Delta, 2\Delta$ and 3Δ), F_{max}, and the maximum of the errors when predicting the averaged discrete behaviour with the continuum models. We notice that there is a trend in all cases, but it becomes noisier as the distance between the lattice sites increases. Additionally, we look at a measure of how correlations are spreading, $F_{AA}(\Delta) - F_{AA}(2\Delta)$, in Figure 6.21(d), observing that this provides the best relationship between its maximum and that of the continuum error predictions. Thus, we can be reasonably sure when using this measure to determine which model is best suited to an experimental system. Our results show that it is the correlations between nearby sites, and how these correlations spread, that has the most significant effect on the continuum predictions.

6.4.2 Two-species system

Turning to the two-species case, we now consider both the auto-correlation functions, F_{AA} and F_{BB}, and the cross-correlation functions, F_{AB}. For the auto-correlation functions we take the overall maximum, F_{max}, to be the highest of the maxima of

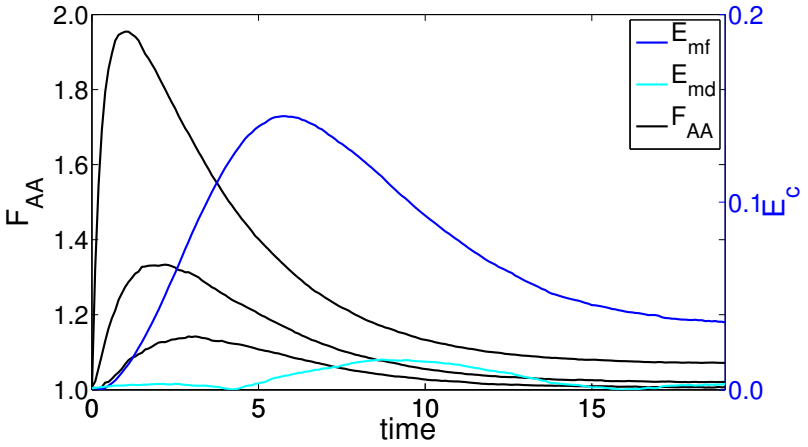
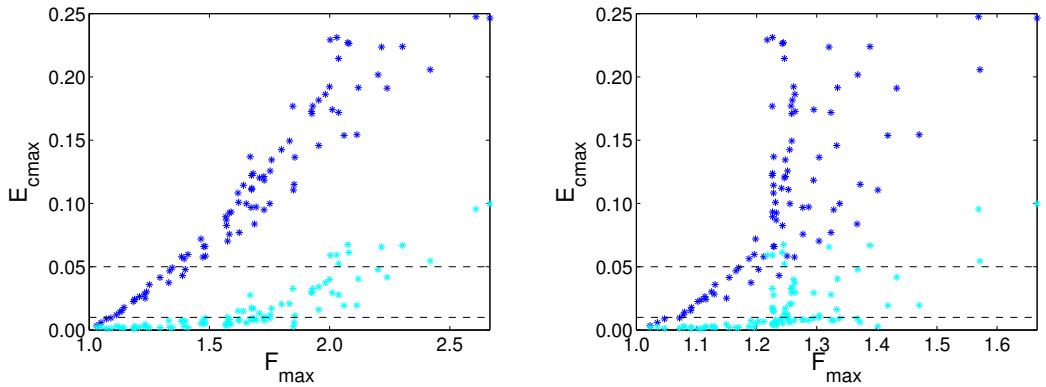


Figure 6.20: The maxima of the correlation functions (black, decreasing as the distance increases from Δ to 3Δ in steps of Δ) does not occur at the same point as the maxima of the errors when predicting the averaged discrete behaviour using the MFA (blue) or correlations model (cyan). Thus we again cannot rely on data collected at a single point in time to determine which model is most suited to a problem. In this case, $P_m = P_p = 1, P_d = 0.4$.

F_{AA} and F_{BB}. There is again a general trend between the maxima of the correlation functions and the continuum prediction errors, although there is quite a large spread in the points, which increases as the distance between the lattice sites increases (Figure 6.22). In the one-species case, we find that $F_{AA}(\Delta) - F_{AA}(2\Delta)$ gives the best relationship between the maxima, therefore we test this in the two species case for the auto-correlation and cross-correlation functions. We find that, as for the other clustering methods we have tested, there is a larger spread in the points in the two-species case thus we cannot confidently use it to determine which model is the best choice for an experimental system.

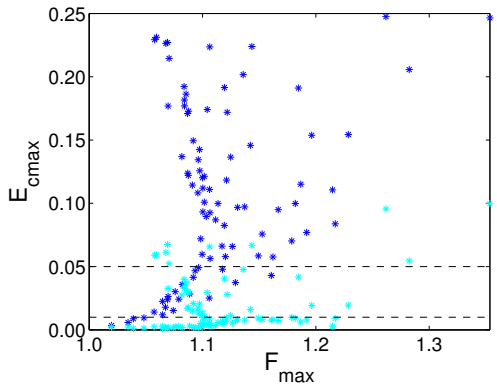
6.5 Comparing the clustering measures

The three different methods presented in this chapter all attempt to measure how far our system is from the case where agents are completely randomly distributed. The ACN method considers the number of occupied sites around an agent: if agents

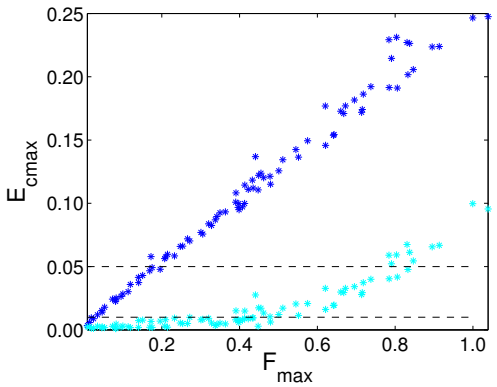


(a) $F_{AA}(\Delta)$

(b) $F_{AA}(2\Delta)$

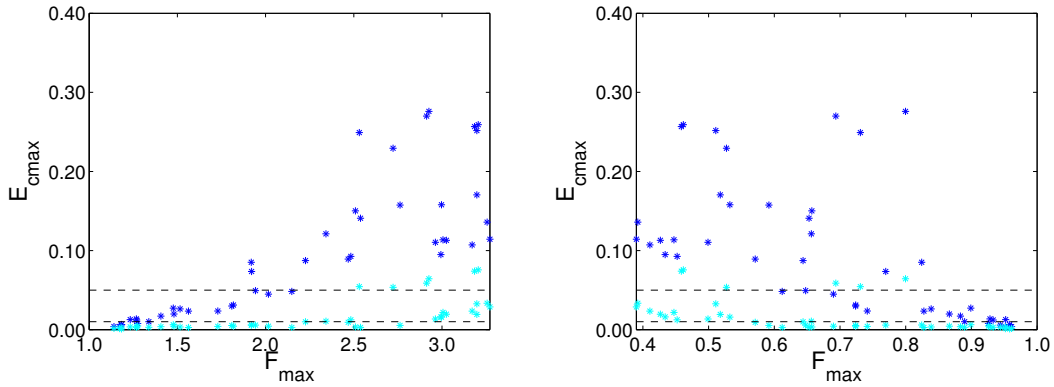


(c) $F_{AA}(3\Delta)$



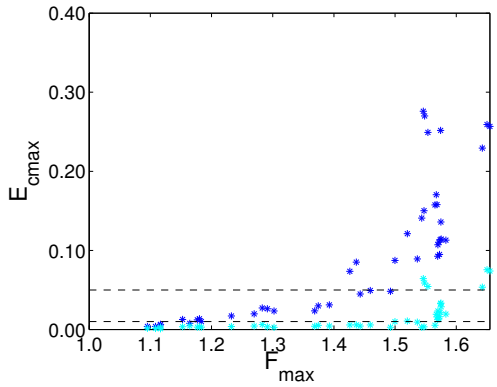
(d) $F_{AA}(\Delta) - F_{AA}(2\Delta)$

Figure 6.21: The maxima of the correlation functions and continuum prediction errors, with the mean-field in blue and correlations in cyan, are related. We see that the relationship becomes noisier as the distance between the lattice sites increases, (a)-(c). The most consistent trend is seen when using the difference between the maxima at distances of Δ and 2Δ as our measure, (d). Thus this is the best correlations measure to use to determine which model to choose for a system.

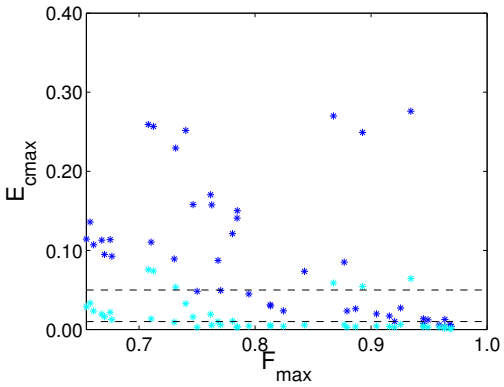


(a) auto-correlation at Δ

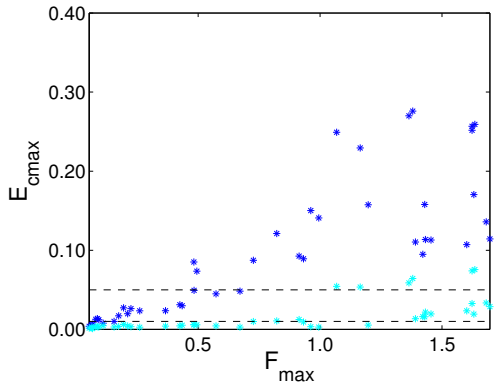
(b) cross-correlation at Δ



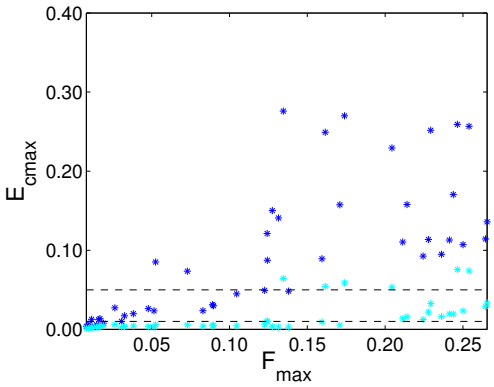
(c) auto-correlation at 2Δ



(d) cross-correlation at 2Δ



(e) auto-correlation difference



(f) cross-correlation difference

Figure 6.22: In the two-species case, we find that the maxima of the correlation functions and continuum prediction errors (mean-field in blue, correlations in cyan) are not as closely related as in the one-species case. We examine the maxima of our cross-correlation and auto-correlation functions at distances of Δ and 2Δ, as well as their difference.

tend to cluster, we expect there to be more occupied sites around a neighbour than in the random case. The spatial index method divides the domain into boxes and evaluates the variance in the number of agents per box: more clustering leads to a greater variance. The correlations method measures the correlation functions, which will deviate from unity if the agents tend to cluster.

We found that, for all our methods, we had to consider the evolution of the system over time. We could not merely compare model predictions against data collected at a single point in time and use this as an indication of the system as a whole. This is because there is a delay between systems reaching the maxima of our clustering indices and of the errors in our model density predictions. This delay is to be expected as the build-up of correlations will then cause a decrease in the possible proliferation events as the agents will have less access to unoccupied sites. We find, from the correlations method, that it is very local effects that are important: the relationship between nearest neighbour correlations and the density errors is more consistent than that for larger distances. We also note, by using the difference between correlations at the first and second distance, that the way in which correlations spread is closely related to the error in the density predictions. These observations from the correlations model explain the relative success of the ACN and SI methods. We saw, in Section 6.2, that there was a fairly consistent relationship between the maxima of our density errors and E_k. Using the ACN, we are considering the occupancies of nearest-neighbour sites, thus are limiting our results to local correlations, which we know to be quite strongly related to the errors. However, in the SI method, we are considering larger regions of the domain, thus taking into account longer range correlations. Thus, whilst the SI method can provide an indication as to whether the system is close to a random state, it does not provide as strong a relationship with the density prediction errors.

In the one-species case, the maximum values of all three methods are correlated

Method	1% MFA	5% MFA	1% Corr PDE	5% Corr PDE
ACN	0.07	0.28	0.52	0.73
SI($\times 10^{-3}$)	0.60	n/a	n/a	n/a
$\boldsymbol{F_{max}}$	0.03	0.20	0.41	0.83

Table 6.1: The maximum value of our methods can be used to indicate which model may be suitable for given experimental data. Here we give the values of our measures at the point at which the error in the continuum approximations to the averaged discrete results is expected to reach a threshold value. F_{max} is the difference in the maxima of the correlation functions at distances of Δ and 2Δ.

with the maximum of the mean-field error, allowing us to be able to reasonably estimate cutoff values for the 1% MFA threshold. However, for other thresholds we considered, the SI method no longer provides a consistent relationship between the maxima thus is best avoided as a method for determining whether or not to use the continuum models in these cases. The thresholds for the 1% and 5% error cutoffs in the one-species case, where they can be determined, can be seen in Table 6.1.

Of course, we have tested our methods on completely theoretical data, all of which have been produced by on-lattice models. The SI method can be easily applied to experimental data as we would still divide the domain and then detect and count the agents in each bin, although we would need to account for cells lying on the boundaries. However, it would not be as straightforward with the ACN or correlations methods as these lend themselves more easily to on-lattice systems. Instead of examining the eight neighbours for the ACN, we would have to consider the number of cells within a specified distance from the centre of a chosen cell. Similary, for the correlations method, we would count the number of pairs whose centres are separated by less than a specified distance. Moreover, in our theoretical cases, we average over one hundred discrete simulations. Most experiments will not have this many repeats, thus we also need to take this into account when using our models on experimental data as we expect there to be a larger variation in the results.

Despite their success in the one-species case, we found that our methods did not simply extend to two-species systems. This could in part be because there are more

degrees of freedom in two-species models as they have twice as many parameters, thus there will be more variability. Moreover, the behaviour of one species will affect the other. We can obtain some indication of the 1% threshold for the MFA, but for all other threshold values we cannot determine any reasonable measure. As we saw, even if a continuum model is suitable for modelling a species alone, it may no longer be a suitable approximation if that species is mixed with another species. Thus, we must proceed with extreme caution when selecting models for multi-species systems.

6.6 Conclusions

In this chapter we have presented and tested three methods to provide a measure of the clustering of a system beginning from uniform initial conditions. The ACN compares the number of occupied sites around an agent to that expected in the random case. We find that this method can, for a one species system, provide a reasonable measure as to the maximum error we could expect when using the mean-field or correlations models to predict the averaged discrete behaviour. However, when used for the two species case, we find the ACN method no longer provides a clear measure. Using the SI we divide our domain into sections and examine the scaled variance of the number of agents in each section. We find that the index is dependent on the number of bins into which we divide our domain, thus making it harder to define a measure than with the ACN. Moreover, the trend is not as clear in the one-species case as it is with the ACN. However, in the two-species case, whilst the relationship is still noisy, we can at least use it as a cut-off for the 1% threshold in the MFA. Finally, we used our correlations functions as a method for determining the level of clustering in our system. We found that there was again a trend between the maxima of our correlation functions and the error in our continuum density predictions. The best correlations measure was the difference between the maxima of the correlations at distances of

Δ and 2Δ. Using this measure, we could confidently determine the 1% and 5% thresholds for the mean-field and continuum models. However, the measure once again did not extend to the two-species case. Table 6.1 indicates the threshold values for each of the methods, thus these can be used as a rough guide to determine which method to choose for a one-species system beginning from uniform initial conditions.

For all our methods we found that following the system's behaviour over time is essential as there is a delay between the maximum level of clustering and the maximum errors we observe in our continuum predictions. Thus, whichever method is used to determine model choice for an experimental system, we must be sure to track the evolution of the clustering rather than evaluating only a single point in time. By comparing our methods, we realise that the levels of very local clustering are related to the errors observed in the continuum models at predicting the discrete behaviour. In the case of our pair-wise correlations model, this can be understood by remembering that the KSA breaks down when the sites in question are very close together (Figure 2.6). Thus it will be local clustering that contributes to the break down of the model, and this provides the best estimate of the error.

Thus far, in this thesis, we have discussed models that go beyond the mean-field, with a primary focus on correlations models, and in this chapter we have evaluated methods for determining which models can be applied to a specific experimental set-up. In the next chapter, we turn to the conclusions we have drawn and discuss the future directions that have been opened up from the work contained in this thesis.

Chapter 7

Conclusion

In this thesis we have demonstrated that mean-field methods can successfully be used to relate individual-level and population-level models if the rates of migration are relatively high when compared to those of proliferation. In cases where the MFA no longer holds, we have provided an alternative model which incorporates spatial correlations, thus including information about the spatial relationships between site occupancies. Our work has provided a simplification to an existing one-species model with homogeneous initial conditions, as well as further developing this model to account for multiple species. We have also examined the non-homogeneous initial conditions case, comparing a correlations model to the OHA across a range of parameters and time-scales. Moreover, we have investigated methods evaluating the level of clustering which can be used in some situations to determine which model is appropriate for given experimental data.

In Chapter 2, we presented an existing model which incorporates correlations into a one-species system undergoing birth, migration and death on a regular lattice in 1D, 2D and 3D with uniform initial conditions [Baker and Simpson, 2010]. We demonstrated an improvement upon the MFA at predicting the averaged discrete behaviour in a range of parameter space. However, we found that this improved model was not

straightforward to implement in 2D and 3D due to the irregular lattice site distances in the correlation function ODEs, limiting its tractibility and making it challenging to extend to more realistic systems. We saw, in Chapter 3, how the system is even more prohibitively complex when we have non-uniform initial conditions as we can no longer assume translational invariance, which greatly simplified our model in the case with uniform initial conditions. For the non-homogeneous initial conditions, we examined an existing correlations model [Simpson and Baker, 2011] in more detail considering both the transient and asymptotic behaviour. We found that the correlations model improved upon the mean-field at approximating the averaged discrete behaviour in the transient region, but the asymptotic behaviour was identical. We noticed that the correlations and mean-field models perform well when rates of proliferation are low relative to those of migration, but begin to break down as the rates of proliferation become large relative to those of migration. Thus these models are not suitable in regions of parameter space where proliferation is significantly more rapid than migration. In addition to the correlations model for inhomogeneous initial conditions, we also extended the model of Callaghan et al. [2006], the OHA. This model uses series expansions to determine the asymptotic speed of a moving front, assuming there is only one unoccupied site behind the front. The OHA cannot be used for explaining the transient behaviour, but provides an excellent approximation of the asymptotic front speed when P_p/P_m is large. However, it breaks down when P_p/P_m is small as this frequently leads to there being more than one unoccupied site behind the moving front. Comparing the MFA, OHA and correlations model across a range of parameter space and time-scales enabled us to examine which methods might be most suitable for specific experimental conditions.

In Chapter 4, we simplified the system presented in Chapter 2 by representing the system of correlation ODEs as a PDE. This model is easier to implement, particularly in situations where long-range correlations may be important, making it more

amenable to various extensions. Moreover, our PDE is in the form of a reaction-diffusion equation, allowing us to examine the roles that migration and proliferation play in the correlation functions. Additionally, previous work [Baker and Simpson, 2010] focussed attention only on the parameter space wherein $P_p/P_m \leq 1$. Having noticed the break down of the correlations model in these regions of parameter space with non-uniform conditions in Chapter 3, we extended their investigations by considering cases where $P_p/P_m > 1$. Generally, the correlations model is found to still provide a valid approximation to the averaged discrete behaviour. However, there are cases, particularly where the death rate is high relative to that of proliferation, when the discrete simulations predict that the species will go to extinction, the absorbing state of the system. This behaviour is not always reflected in the correlations model, thus we must proceed with caution when using our model in these regions of parameter space.

Our PDE model presented in Chapter 4 allows for extensions, thus enabling us to, for example, more readily examine systems with multiple species. In Chapter 5, we presented a correlations PDE model for the N-species case, and showed results for $N = 2$. We found that the MFA was again a valid approximation when P_p/P_m was very low, but otherwise was not a good approximation to the averaged discrete behaviour. Not only is the MFA frequently unable to accurately predict the transient behaviour, but there are also cases where it is erroneous in predicting the asymptotic results. For example, the averaged discrete behaviour may indicate that a given species will dominate at long times, with the other species going to extinction, but the MFA makes the opposite prediction. This dramatic difference in results is usually rectified by incorporating spatial correlations into our model. Naturally, the correlations model has its own shortcomings and is not a valid approximation in all scenarios either. Particularly when migration rates are low, and we are near a switching point, we sometimes found our correlations model also making inaccurate predictions. It is,

however, always seen to be an improvement upon the MFA.

Our model results demonstrated that choosing the wrong model framework for a biological application could have serious consequences. Choosing which model is best is not trivial, and in Chapter 6 we discussed some methods which can potentially be used under certain conditions. We compared these measures to the errors observed when predicting the averaged discrete behaviour with our two continuum models, the MFA and correlations PDE models. The methods we chose were the ACN (where we compare the number of occupied sites around an agent to that expected in the random case) [Simpson et al., 2014], the SI (where we consider the variance in the number of agents in divided regions of the domain) [Binder and Landman, 2011], and the correlation functions (where we calculate the correlation functions directly from the data) [Treloar et al., 2014]. All of our measures vary with time, and this variation does not match that of the errors in our continuum predictions. Thus we cannot simply consider one point in time, but must take into account the behaviour over a range of time. We determine if there is a relationship between the maxima of the measures and errors in both the one-species and two-species cases. We find that the ACN and correlations methods can successfully be used in the one-species case. The SI is not as useful as it depends on the number of regions into which we divide our domain, and the trend between the maxima is also not as clear. Despite their success in the one-species case, we find that none of our methods are easily extended to the two-species case, thus different methods will need to be developed for multi-species systems.

In this thesis, we have made progress at going beyond mean-field models in a number of situations, and have identified some important next steps, some of which we shall now discuss.

7.1 Future directions

There are many directions that can be taken from the work we have done thus far. One of the most natural directions would be to derive and test a PDE, similar to that in Chapter 4, for the correlations model with non-uniform initial conditions. This would enable us to more readily look at systems with non-uniform initial conditions in higher dimensions, which are more realistic for biological systems. For the case with homogeneous initial conditions, our PDE depends on time and the radial distance between lattice sites. As we observed in Chapter 3, without the assumption of translational invariance the correlation functions are no longer dependent on just the radial distance, but on the specific sites the two agents in question occupy. Thus, our PDE will now have an extra variable: it will depend on time, the position of the specific site, and the position of the correlation site. A PDE has been derived that accounts for only nearest-neighbour correlations, specifically for the case considering the effect of adhesion and contact interactions [Ascolani et al., 2013]. It would be sensible to begin with a nearest-neighbour PDE in our case as well, before extending it to longer-range correlations.

Once a PDE is obtained for the case with non-uniform initial conditions, it will be more straightforward to extend this system to account for multiple species, using the method in Chapter 5. This could be used to rigorously derive continuum models for a number of biological situations. Classical models of tumour invasion use coupled reaction-diffusion equations to represent healthy tissue and a tumour [Gatenby and Gawlinski, 1996; Gatenby et al., 2006]. Rigorously deriving equations for these phenomena could give more accurate insights into the biological processes at hand. Whilst this thesis has focused primarily on applications in cell biology, these methods are equally applicable in other areas of mathematical biology and ecology, as seen in Chapter 1. Being able to derive multi-species models for non-uniform initial conditions could have many other applications, such as modelling invasive species in an

ecological setting [Gurnell, 1987; MacKinnon, 1978].

Off-lattice models provide a more realistic approximation of biological phenomena [Dyson et al., 2012], thus we could develop methods similar to those presented in this thesis that would be suitable for off-lattice IBMs. Off-lattice moment dynamics models for plants have been developed [Law et al., 2003], which use continuous dispersal and competition kernels. However, as of yet, no off-lattice moment dynamics models exist for cellular processes. Thus, future work could involve deriving a correlations PDE model for an off-lattice cellular system. Moreover, on-lattice models are not restricted to those that are square: hexagonal lattices are often used and have been included in other correlations models [Ascolani et al., 2013]. We could extend our correlations model in a similar fashion to account for hexagonal lattices.

We have only examined our methods for a specific set of rules in our IBM. For the single-species case, proliferation is nearest-neighbour only, and death is purely random. Proliferation is not necessarily confined to neighbouring sites, particularly in ecology; plant seeds are often carried long distances by animals [Sorenson, 1986] and wind dispersal [Gurevitch et al., 2006]. Thus we could modify our proliferation rules to account for this. This would mean that the density evolution will no longer depend solely on nearest-neighbour correlation terms. Additionally, death in biological systems is often density-dependent [Greenhalgh, 1990; Guckenheimer et al., 1977]. Many biological individuals in close proximity can have a negative effect on one another as they may not have optimal access to resources. Thus we could modify our death events to depend on the number of surrounding agents. This will most likely cause our death terms to depend on the correlation functions. At the moment, proliferative and movement events are aborted if the target site is occupied. However, in biology, proliferating cells do not necessarily require unoccupied space near them, but would rather just push nearby cells away to make room for the daughter cell [Alpkvist et al., 2006; Laspidou et al., 2010]. We could implement a different

mechanism in our model, whereby the moving agent, or daughter cell, occupies the site and 'shoves' the current occupant elsewhere. This could decrease the need for including correlations in our model as we will not be relying on nearby empty spaces for successful proliferation to occur.

We could also modify the rules in our multi-species case. In the model presented in Chapter 5, the species are not directly interacting. We could incorporate specific interactions, such as competition for resources or predation. We could include some different mechanisms in the same way as we could in the single-species case; for example, events could be dependent on the number of agents of a similar or different species in the surrounding sites. Allelopathy, the effect of an organism on another by secretion of chemical compounds into the environment [Rice, 1984; Willis, 2007], plays a significant role in the interactions of bacterial species. For example, bacteriocins are compounds produced by a bacterial strain, to which the species itself is immune. However, the compounds can be harmful to other strains [James et al., 1992; Riley and Chavan, 2006]. This spiteful behaviour comes at a cost to the toxin-producing species; therefore, there must be some cost-benefit trade-off at play in these systems. We could extend our two-species model in Chapter 5 to account for bacteriocin production, enabling us to examine the relative benefits of toxin production in different parameter regimes.

We found, in Chapter 6, that our clustering measures generally provided a good measure of which model may be suited to a one-species problem. However, these measures did not apply so well for a two-species system, thus we need to investigate a more generally applicable measure. Moreover, our methods only apply to systems beginning from uniform initial conditions. Similar methods for non-uniform initial conditions would also be useful. We saw, in Chapter 3, that the MFA and correlations model do not perform as well when there are fewer holes behind the front, thus we could perhaps try using the average number of holes behind the front as a measure

to determine which model is suited to experimental data of an invading front.

The work contained in this thesis has been completely theoretical. Testing these methods on experimental systems can determine if our observations thus far have been accurate representations of biological systems. Work has been done on comparing the performance of the mean-field and correlations model when estimating parameters from experimental data beginning from random uniform initial conditions [Simpson et al., 2013a]. We could examine our correlations and OHA models for travelling fronts in a similar way. Additionally, it would be interesting to find competing populations in which one drives the other to extinction, and determine if there are any populations in which the correlations model predicts the correct behaviour whilst the MFA predicts the opposite, as we saw in Chapter 5. Additionally, the clustering measures we discussed in the previous chapter were only tested with on-lattice theoretical data. These would need to be tested in experimental systems to determine if they are readily applicable to experimentally generated data.

Being able to rigorously derive population models directly from individual interactions is crucial when modelling biological phenomena. In this thesis we have presented methods which have made some progress in this challenging area of research. Our work has opened up other potential avenues of inquiry, of interest not only to theoretical researchers but also to those more concerned with biological applications.

Bibliography

Alarcón, T., Byrne, H. M., and Maini, P. K. (2003). A cellular automaton model for tumour growth in inhomogeneous environment. *J. Theor. Biol.*, 225(2):257–274.

Allred, D. C. (2010). Ductal carcinoma in situ: Terminology, classification and natural history. *J. Nat. Cancer Inst. Monographs*, 41:134–138.

Alpkvist, E., Picioreanu, C., van Loosdrecht, M. C. M., and Heyden, A. (2006). Three-dimensional biofilm model with individual cells and continuum EPS matrix. *Biotechnol. Bioeng.*, 94(5):961–979.

Anderson, R. M. and May, R. M. (1979). Population biology of infectious diseases: Part I. *Nature*, 280 (5721):361–7.

Ascolani, G., Badoual, M., and Deroulers, C. (2013). Exclusion processes: short range correlations induced by adhesion and contact interactions. *Phys Rev. E*, 87:012792.

Bailey, J. E. and Ollis, D. F. (1986). *Biochemical Engineering Fundamentals*. McGraw Hill, New York.

Baker, R. E. and Simpson, M. J. (2010). Correcting mean-field approximations for birth-death-movement processes. *Phys. Rev. E.*, 82:041905.

Barrandon, Y. and Green, H. (1987). Cell migration is essential for sustained growth of keratinocyte colonies: the roles of transforming growth factor-α and epidermal growth factor. *Cell*, 50(7):1131–1137.

Barraquand, F. and Murrell, D. J. (2013). Scaling up predator-prey dynamics using spatial moment equations. *Method Ecol Evol*, 4:276–289.

Binder, B. J. and Landman, K. A. (2011). Quantifying evenly distributed states in exclusion and nonexclusion processes. *Phys. Rev. E*, 83:041914.

Birch, D. A. and Young, W. R. (2006). A master equation for a spatial population model with pair interactions. *Theor. Pop. Biol.*, 70:26–42.

Bolker, B. M. and Pacala, S. W. (1997). Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theor. Pop. Biol.*, 52:179–197.

Bolker, B. M. and Pacala, S. W. (1999). Spatial moment equations for plant competition: understanding spatial strategies and the advantages of short dispersal. *Am. Nat.*, 153(6):575–602.

Bolker, B. M., Pacala, S. W., and Neuhauser, C. (2003). Spatial dynamics in model plant communities: what do we really know? *Am. Nat.*, 162(2):135–148.

Boots, M. and Sasaki, A. (1999). 'Small worlds' and the evolution of virulence: infection occurs locally and at a distance. *Proc. R. Soc. Lond. B*, 266:1933–1938.

Brauer, F. A. and Castillo-Chavez, C. (2001). *Mathematical Models in Population Biology and Epidemiology*. Springer.

Byrne, H. M., Alarcón, T., Owen, M. R., Webb, S. D., and Maini, P. K. (2006). Modelling aspects of cancer dynamics: a review. *Phil. Trans. R. Soc. A*, 364(1843):1563–1578.

Cabrero, A., Fernandez, S., Fernando, M., and Garcia, J. (1998). Effects of copper and zinc on the activated sludge bacteria growth kinetics. *Water Res.*, 32(5):1355–1362.

Callaghan, T., Khain, E., Sander, L. M., and Ziff, R. M. (2006). A stochastic model for wound healing. *J. Stat. Phys.*, 122(5):909–923.

Chao, L. and Levin, B. R. (1981). Structured habitats and the evolution of anti-competitor toxins in bacteria. *Proc. Nat. Acad. Sci.*, 78:6324–6328.

Cheng, G., Youssef, B. B., Markenscoff, P., and Zygourakis, K. (2006). Cell population dynamics modulate the rates of tissue growth processes. *Biophys. J.*, 90(3):713–724.

Codling, E. A., Plank, M. J., and Benhamou, S. (2008). Random walk models in biology. *J. R. Soc. Interface*, 5:813–834.

Cox, J. T. and Schinazi, R. B. (2009). Survival and coexistence for a multitype contact process. *Ann. Probab.*, 3:853–876.

Cuddington, K. M. and Yodzis, P. (2000). Diffusion-limited predator-prey dynamics in euclidean environments: an allometric individual-based model. *Theor. Pop. Biol.*, 58:259–278.

Czaran, T. and Bartha, S. (1992). Spatiotemporal dynamic models of plant populations and communities. *Trends Ecol. Evol.*, 7(2):38–42.

Dale, P., Maini, P. K., and Sherratt, J. A. (1994). Mathematical modeling of corneal epithelial wound healing. *Math. Biosci.*, 124(2):127–147.

Dallon, J., Sherratt, J., Maini, P. K., and Ferguson, M. (2000). Biological implications of a discrete mathematical model for collagen deposition and alignment in dermal wound repair. *Math. Med. Biol.*, 17(4):379–393.

Dangerfield, C. E., Ross, J. V., and Keeling, M. J. (2009). Integrating stochasticity and network structure into an epidemic model. *J. R. Soc. Interface*, 6(38):761–774.

De Roos, A. M., McCauley, E., and Wilson, W. G. (1991). Mobility versus density-limited predator-prey dynamics on different spatial scales. *Proc. R. Soc. Lond. B*, 246:117–122.

Denison, T. A. and Bae, Y. H. (2012). Tumor heterogeneity and its implication for drug delivery. *J. Control Release*, 164(2):187–191.

Deroulers, C., Aubert, M., Badoual, M., and Grammaticos, B. (2009). Modeling tumor cell migration: From microscopic to macroscopic models. *Phys. Rev. E*, 79:031917(1–14).

Dickman, R. (1989). Nonequilibrium lattice models: series analysis of steady states. *J. Stat. Phys.*, 55:997–1026.

Dieckmann, U. and Law, R. (2000). *Relaxation Projections and the Method of Moments*, chapter 21, pages 412–457. Cambridge University Press.

Dobrushin, R. L. (1986). *Mathematical Problems of Statistical Mechanics and Dynamics: A Collection of Surveys*. Springer.

Dodd, P. J. and Ferguson, N. M. (2009). A many-body field theory approach to stochastic models in population biology. *Plos One*, 4(9):e6855.

Dormann, S. and Deutsch, A. (2002). Modeling of self-organized avascular tumor growth with a hybrid cellular automaton. *In silico Biol.*, 2(3):393–406.

Durrett, R. and Levin, S. (1994). The importance of being discrete (and spatial). *Theor. Pop. Biol.*, 46:363–394.

Dyson, L., Maini, P. K., and Baker, R. E. (2012). Macroscopic limits of individual-based models for motile cell populations with volume exclusion. *Phys. Rev. E*, 86:031903.

Ellner, S. P. (2001). Pair approximation for lattice models with multiple interaction scales. *J. Theor. Biol.*, 210:435–447.

Ellner, S. P., Sasaki, A., Haraguchi, Y., and Matsuda, H. (1998). Speed of invasion in lattice population models: pair-edge approximation. *J. Math. Biol.*, 36:469–484.

Ermentrout, G. B. and Edelstein-Keshet, L. (1993). Cellular automata approaches to biological modeling. *J. Theor. Biol.*, 160:97–133.

Ferguson, N. M., Donnelly, C. A., and Anderson, R. M. (2001). The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science*, 292:1155–1160.

Filipe, J. A. N. and Gibson, G. J. (2001). Comparing approximations to spatio-temporal models for epidemics with local spread. *Bull. Math. Biol.*, 63:603–624.

Filipe, J. A. N. and Maule, M. M. (2003). Analytical methods for predicting the behaviour of population models with general spatial interactions. *Math. Biosci.*, 183:15–35.

Fisher, R. A. (1937). The wave of advance of advantageous genes. *Ann. Eugenics*, 7:353–369.

Galle, J., Aust, G., Schaller, G., Beyer, T., and Drasdo, D. (2006). Individual cell-based models of the spatio-temporal organization of multicellular systems - acheivements and limitations. *Cytometry Part A*, 69A(7):704–710.

Gatenby, R. A. and Gawlinski, E. T. (1996). A reaction-diffusion model of cancer invasion. *Cancer Res.*, 56(24):5745.

Gatenby, R. A., Gawlinski, E. T., Gmitro, A. F., Kaylor, B., and Gillies, R. J. (2006). Acid-mediated tumor invasion: a multidisciplinary study. *Cancer Res.*, 66(10):5216.

Gatenby, R. A., Smallbone, K., Maini, P. K., Rose, F., Averill, J., Nagle, R. B., Worrall, L., and Gillies, R. J. (2007). Cellular adaptations to hypoxia and acidosis during somatic evolution of breast cancer. *Brit. J. Cancer*, 97(5):646–653.

Gerlinger, M., Rowan, A. J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., Varela, I., Phillimore, B., Begum, S., McDonald, N. Q., Butler, A., Jones, D., Raine, K., Latimer, C., Santos, C. R., Nohadani, M., Eklund, A. C., Spencer-Dene, B., Clark, G., Pickering, L., Stamp, G., Gore, M., Szallasi, Z., Downward, J., Futreal, P. A., and Swanton, C. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *New Engl. J. Med.*, 366(10):883–892.

Gilbert, S. F. (2006). *Development Biology*. Sinauer Associates Inc., Sunderland, MA.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361.

Grandy, W. T. J. (1987). *Foundations of Statistical Mechanics: Volume 1: Equilibrium Theory*. Springer.

Greenhalgh, D. (1990). An epidemic model with a density-dependent death rate. *Math. Med. Biol.*, 7(1):1–26.

Grimm, V. (1999). Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? *Ecol. Model.*, 115(2-3):129–148.

Guckenheimer, J., Oster, G., and Ipaktchi, A. (1977). The dynamics of density dependent population models. *J. Math. Biol.*, 4:101–147.

Gurevitch, J., Scheiner, S. M., and Fox, G. A. (2006). *Plant Ecology, 2nd Edition*. Sinnauer Associates, Inc., Massachusetts.

Gurnell, J. (1987). *The Natural History of Squirrels*. London: Christopher Helm.

Harada, Y. (1999). Short- vs. long-range disperser: the evolutionarily stable allocation in a lattice-structured habitat. *J. Theor. Biol.*, 201:171–187.

Haraguchi, Y. and Sasaki, A. (2000). The evolution of parasite virulence and transmission rate in a spatially structured population. *J. Theor. Biol.*, 203:85–96.

Harris, T. E. (1974). Contact interactions on a lattice. *Ann. Probab.*, 2(6):969–988.

James, R., Lazdunski, C., and Pattus, F. (1992). *Bacteriocins, Microcins and Lantibiotics*. Springer-Verlag.

Jensen, I. and Dickman, R. (1993). Time-dependent perturbation theory for diffusive non-equilibrium lattice models. *J. Phys. A: Math. Gen.*, 26:L151–L157.

Johnston, S. T., Simpson, M. J., and Baker, R. E. (2012). Mean-field descriptions of collective migration with strong adhesion. *Phys. Rev. E*, 85:051922.

Johnston, S. T., Simpson, M. J., and Plank, M. J. (2013). Lattice-free descriptions of collective motion with crowding and adhesion. *Phys. Rev. E*, 88:062720.

Kacena, M. A., Merrell, G. A., Manfredi, B., Smith, E. E., Klaus, D. M., and Todd, P. (1999). Bacterial growth in space flight: logistic growth curve parameters for *Escherichia coli* and *Bacillus subtilis*. *Appl. Microbiol. Biotechnol.*, 51:229–234.

Kamo, M., Sasaki, A., and Boots, M. (2007). The role of trade-off shapes in the evolution of parasites in spatial host populations: an approximate analytical approach. *J. Theor. Biol.*, 244:588–596.

Karrer, B. and Newman, M. E. J. (2010). A message passing approach for general epidemic models. *Phys. Rev. E*, 82:016101.

Keeling, M. J. (1999a). Correlation equations for endemic diseases: externally imposed and internally generated heterogeneity. *Proc. R. Soc. Lond. B*, 266:953–960.

Keeling, M. J. (1999b). The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. Lond. B*, 266:859–867.

Keeling, M. J. (2000). Metapopulation moments: coupling, stochasticity and persistence. *J. Anim. Ecol.*, 69:725–736.

Keeling, M. J. and Eames, K. T. D. (2005). Networks and epidemic models. *J. R. Soc. Interface*, 2:295–307.

Keeling, M. J., Rand, D. A., and Morris, A. J. (1997). Correlation models for childhood epidemics. *Proc. R. Soc. Lond. B*, 264:1149–1156.

Kermack, W. and McKendrick, A. (1991a). Contributions to the mathematical theory of epidemics - I. *Bull. Math. Biol.*, 53(1-2):33–55.

Kermack, W. and McKendrick, A. (1991b). Contributions to the mathematical theory of epidemics - II. the problem of endemicity. *Bull. Math. Biol.*, 53(1-2):57–87.

Kermack, W. and McKendrick, A. (1991c). Contributions to the mathematical theory of epidemics - III. further studies of the problem of endemicity. *Bull. Math. Biol.*, 53(1-2):89–118.

Kessler, D. A. and Sander, L. M. (2009). Fluctuations and dispersal rates in population dynamics. *Phys. Rev. E*, 80:041907.

Khain, E., Lin, Y. T., and Sander, L. M. (2011). Fluctuations and stability in front propagation. *EPL*, 93:28001.

Kim, Y., Stolarska, M. A., and Othmer, H. G. (2007). A hybrid model for tumor spheroid growth *in vitro*: I. theoretical development and early results. *Math. Methods Appl. Sci.*, 17:1773–98.

Kirkwood, J. G. (1935). Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 3:300–314.

Kirkwood, J. G. and Boggs, E. M. (1942). The radial distribution function in liquids. *J. Chem. Phys.*, 10:394–403.

Kolmogorov, A., Petrovsky, I., and Piscounov, N. (1937). Study of the diffusion equation with growth of the quantity of matter and its application to a biological problem. *Bull. State Univ. Mos.*, pages 1–25.

Laspidou, C. S., Kungolos, A., and Samaras, P. (2010). Cellular-automata and individual-based approaches for the modeling of biofilm structures: pros and cons. *Desalination*, 250(1):390–394.

Law, R., Murrell, D. J., and Dieckmann, U. (2003). Population growth in space and time: spatial logistic equations. *Ecology*, 84(1):252–262.

Lewis, M. A. (2000). Spread rate for a nonlinear stochastic invasion. *J. Math. Biol.*, 41:430–454.

Liang, C.-C., Park, A. Y., and Guan, J.-L. (2007). *In vitro* scratch assay: a convenient and inexpensive method for analysis of cell migration *in vitro*. *Nat. Protoc.*, 2(2):329–333.

Liggett, T. M. (1999). *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*. Springer.

Liu, R. T., Liaw, S. S., and Maini, P. K. (2006). Two-stage turing model for generating pigment patterns on the leopard and the jaguar. *Phys. Rev. E*, 74:011914.

Lotka, A. J. (1925). *Elements of Physical Biology*. Williams and Wilkins.

MacKinnon, K. (1978). Competition between red and grey squirrels. *Mamm. Rev.*, 8:185–190.

Mai, J., Kuzovkov, V. N., and von Niessen, W. (1993). A theoretical stochastic model for the a+$1/2b_2 \rightarrow 0$ reaction. *J. Chem. Phys.*, 98:10017–10025.

Mai, J., Kuzovkov, V. N., and von Niessen W. (1994). A general stochastic model for the description of surface reaction systems. *Physica A*, 203:298–315.

Maini, P. K., McElwain, D. L. S., and Leavesley, D. I. (2004a). Traveling wave model to interpret a wound-healing cell migration assay for human peritoneal mesothelial cells. *Tissue Eng.*, 10:475–482.

Maini, P. K., McElwain, D. L. S., and Leavesley, D. I. (2004b). Travelling waves in a wound healing assay. *Appl. Math. Model.*, 17:575–580.

Mani, S., Zygourakis, K., and Markenscoff, P. (2002). Modeling wound healing: effects of cell heterogeneity and tissue structure. In *Proceedings of the Second Joint EMBS/BMES Conference, Houston, TX, USA*.

Martin, P. C. (1968). *Measurements and Correlation Functions*. CRC Press.

Marusyk, A., Almendro, V., and Polyak, K. (2012). Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer*, 12:323–334.

Matsuda, H., Ogita, N., Sasaki, A., and Sato, K. (1992). Statistical mechanics of population. *Prog. Theor. Phys.*, 88(6):1035–1049.

May, R. M. (1975). *Stability and Complexity in Model Ecosystems*. Princeton University Press, Princeton, USA.

May, R. M. and Anderson, R. M. (1979). Population biology of infectious diseases: Part II. *Nature*, 280 (5722):455–61.

McLennan, R., Dyson, L., Prather, K. W., Morrison, J. A., Baker, R. E., Maini, P. K., and Kulesa, P. M. (2012). Multiscale mechanisms of cell migration during development: theory and experiment. *Development*, 139:2935–2944.

Messinger, S. M. and Ostling, A. (2009). The consequences of spatial structure for the evolution of pathogen transmission rate and virulence. *Am. Nat.*, 174:441–454.

Mi, Q., Riviere, B., Clermont, G., Steed, D. L., and Vodovotz, Y. (2007). Agent-based model of inflammation and wound healing: insights into diabetic foot ulcer pathology and the role of transforming growth factor-β1. *Wound Repair. Regen.*, 15(5):671–682.

Murray, J. D. (2002). *Mathematical Biology I: An Introduction*. Springer.

Murrell, D. J. (2005). Local spatial structure and predator-prey dynamics: counter-intuitive effects of prey enrichment. *Am. Nat.*, 166(3):354–367.

Murrell, D. J. and Law, R. (2000). Beetles in fragmented woodlands: a formal framework for dynamics of movement in ecological landscapes. *J. Anim. Ecol.*, 69:471–483.

Okubo, A., Maini, P. K., Williamson, M. H., and Murray, J. D. (1989). On the spatial spread of the grey squirrel in Britain. *Proc. R. Soc. Lond. B*, 238(1291):113–125.

Othmer, H. G., Dunbar, S. R., and Alt, W. (1988). Models of dispersal in biological systems. *J. Math. Biol.*, 26(3):263–298.

Ovaskainen, O., Sato, K., Bascompte, J., and Hanski, I. (2002). Metapopulation models for extinction threshold in spatially correlated landscapes. *J. Theor. Biol.*, 215:95–108.

Owen, M. R., Alarcón, T., Maini, P. K., and Byrne, H. M. (2009). Angiogenesis and vascular remodelling in normal and cancerous tissues. *J. Math. Biol.*, 58:689–721.

Parham, P. E. and Ferguson, N. M. (2006). Space and contact networks: capturing the locality of disease transmission. *J. R. Soc. Interface*, 3:483–493.

Pennington, C. J., Hughes, B. D., and Landman, K. A. (2011). Building macroscale models from microscale probabilistic models: A general probabilistic approach for nonlinear diffusion and multispecies phenomena. *Phys. Rev. E*, 84:041120.

Pielou, E. C. (1969). *An Introduction to Mathematical Ecology*. Wiley, New York.

Plank, M. J. and Simpson, M. J. (2012). Models of collective cell behaviour with crowding effects: comparing lattice-based and lattice-free approaches. *J. R. Soc. Interface*, 9:2983–2996.

Plank, M. J. and Simpson, M. J. (2013). Lattice-free models of cell invasion: discrete simulations and travelling waves. *Bull. Math. Biol.*, 75:2150–2166.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press.

Raghib, M., Hill, N. A., and Dieckmann, U. (2011). A multiscale maximum entropy moment closure for locally regulated space-time point process models of population dynamics. *J. Math. Biol.*, 62:605–653.

Rice, E. L. (1984). *Allelopathy*. Academic Press.

Riley, M. A. and Chavan, M. A. (2006). *Bacteriocins: Ecology and Evolution*. Springer.

Rodriguez, L. G., Wu, X., and Guan, J.-L. (2005). *Methods in Molecular Biology, vol.294: Cell Migration: Developmental Methods and Protocols*, chapter 3 Wound-healing assay, pages 23–29. Humana Press Inc., Totowa, NJ.

Sengers, B. G., Please, C. P., and Oreffo, R. O. C. (2007). Experimental characterization and computational modelling of two-dimensional cell spreading for skeletal regeneration. *J. R. Soc. Interface*, 4:1107–1117.

Sengers, B. G., Please, C. P., Taylor, M., and Oreffo, R. O. C. (2009). Experimental-computational evaluation of human bone marrow stromal cell spreading on trabecular bone structures. *Ann. Biomed. Eng.*, 37(6):1165–1176.

Sharkey, K. J. (2011). Deterministic epidemic models on contact networks: Correlations and unbiological terms. *Theor. Pop. Biol.*, 79(4):115–129.

Sharkey, K. J., Fernandez, C., Morgan, K. L., Peeler, E., Thrush, M., Turnbull, J. F., and Bowers, G. B. (2006). Pair-level approximations to the spatio-temporal dynamics of epidemics on asymmetric contact networks. *J. Math. Biol.*, 53:61–85.

Sharkey, K. J., Kiss, I. Z., Wilkinson, R. R., and Simon, P. L. (2013). Exact equations for SIR epidemics on tree graphs. *Bull. Math. Biol.*

Shigesada, N. and Kawasaki, K. (1997). *Biological Invasions: Theory and Practice*. Oxford University Press.

Shirinifard, A., Gens, J. S., Zaitlen, B. L., Nikodem, J. P., Swat, M., and Glazier, J. A. (2009). 3D multi-cell simulation of tumor growth and angiogenesis. *Plos One*, 4(10):e7190.

Silvertown, J., Holtier, S., Johnson, J., and Dale, P. (1992). Cellular automaton models of interspecific competition for space - the effect of pattern on process. *J. Ecol.*, 80:527–534.

Simpson, M. J. and Baker, R. E. (2011). Corrected mean-field models for spatially dependent advection-diffusion-reaction phenomena. *Phys. Rev. E*, 83:051922.

Simpson, M. J., Binder, B. J., Haridas, P., Wood, B. K., Treloar, K. K., McElwain, D. L. S., and Baker, R. E. (2013a). Experimental and modelling investigation of monolayer development with clustering. *Bull. Math. Biol.*, 75:871.

Simpson, M. J., Hughes, B. D., and Landman, K. A. (2009a). Diffusing populations: Ghosts or folks? *Australas. J. Eng. Ed.*, 15(2):59–66.

Simpson, M. J., Landman, K. A., and Hughes, B. D. (2009b). Multi-species simple exclusion processes. *Physica A*, 388:399–406.

Simpson, M. J., Landman, K. A., and Hughes, B. D. (2010a). Cell invasion with proliferation mechanisms motivated by time-lapse data. *Physica A*, 389:3779–3790.

Simpson, M. J., Merrifield, A., Landman, K. A., and Hughes, B. D. (2007a). Simulating invasion with cellular automata: connecting cell-scale and population-scale properties. *Phys Rev. E*, 76:021918.

Simpson, M. J., Sharp, J. A., and Baker, R. E. (2014). Distinguishing between mean-field, moment dynamics and stochastic descriptions of birth-death-movement processes. *Physica A*, 395:236–246.

Simpson, M. J., Towne, C., McElwain, D. L. S., and Upton, Z. (2010b). Migration of breast cancer cells: Understanding the roles of volume exclusion and cell-to-cell adhesion. *Phys. Rev. E*, 82:041901.

Simpson, M. J., Treloar, K. K., Binder, B. J., Haridas, P., Manton, K. J., Leavesley, D. I., McElwain, D. L. S., and Baker, R. E. (2013b). Quantifying the roles of cell motility and cell proliferation in a circular barrier assay. *J. R. Soc. Interface*, 10:20130007.

Simpson, M. J., Zhang, D. C., Mariani, M., Landman, K. A., and Newgreen, D. F. (2007b). Cell proliferation drives neural crest cell invasion of the intestine. *Dev. Biol.*, 302:553–568.

Singer, A. (2004). Maximum entropy formulation of the Kirkwood superposition approximation. *J. Chem. Phys.*, 121(8):3657–3666.

Smallbone, K., Gatenby, R. A., Gillies, R. J., Maini, P. K., and Gavaghan, D. J. (2007). Metabolic changes during carcinogenesis: potential impact on invasiveness. *J. Theor. Biol.*, 244:703–713.

Smolle, J., Smolle-Juettner, F. M., Stettner, H., and Kerl, H. (1992). Relationship of tumor cell motility and morphologic patterns. Part 1. Melanocytic skin tumors. *Am. J. Dermatopathol.*, 14(3):231–7.

Sorenson, A. E. (1986). Seed dispersal by adhesion. *Ann. Rev. Ecol. Syst.*, 17:443–463.

Spratt, J. A., von Fournier, D., Spratt, J. S., and Weber, E. E. (1993). Decelerating growth and human breast cancer. *Cancer*, 71(6):2013–2019.

Stevens, A. and Othmer, H. G. (1997). Aggregation, blowup, and collapse: The ABC's of taxis in reinforced random walks. *SIAM J. Appl. Math.*, 57(4):1044–1081.

Sumpter, D. J. T. and Broomhead, D. S. (2001). Relating individual behaviour to population dynamics. *Proc. R. Soc. Lond. B*, 268:925–932.

Swanson, K. R., Bridge, C., Murray, J. D., and Ellsworth, C. A. J. (2003). Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion. *J. Neur. Sci.*, 216:1–10.

Takenaka, Y., Matsuda, H., and Iwasa, Y. (1997). Competition and evolutionary stability of plants in a spatially structured habitat. *Res. Popul. Ecol.*, 39(1):67–75.

Todaro, G. J. and Green, H. (1963). Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. *J. Cell Bio.*, 17:299–313.

Tomlin, C. J. and Axelrod, J. D. (2007). Biology by numbers: mathematical modelling in developmental biology. *Nat. Rev. Gen.*, 8:331–340.

Treloar, K. K., Simpson, M. J., Binder, B. J., McElwain, D. L. S., and Baker, R. E. (2014). Assessing the role of spatial correlations during collective cell spreading. *Sci. Rep.*, 4:5713.

Tremel, A., Cai, A., Tirtaatmadja, N., Hughes, B. D., Stevens, G. W., Landman, K. A., and O'Connor, A. J. (2009). Cell migration and proliferation during monolayer formation and wound healing. *Chem. Eng. Sci.*, 64(2):247–253.

Tsoularis, A. and Wallace, J. (2002). Analysis of logistic growth models. *Math. Biosci.*, 179(1):21–55.

Turing, M. A. (1952). The chemical basis of morphogenesis. *Phil. Trans. R. Soc. B*, 237:37–72.

Vaidya, V. G. and Alexandro, F. J. (1982). Evaluation of some mathematical models for tumor growth. *Int. J. Bio-Med. Comp.*, 13(1):19–35.

Valster, A., Tran, N. L., Nakada, M., Berens, M. E., Chan, A. Y., and Symons, M. (2005). Cell migration and invasion assays. *Methods*, 37:208–215.

Verhulst, P.-F. (1838). Notice sur la loi que la population poursuit dans son accroissement. *Correspondance Mathematique et Physique*, 10:113–121.

Volterra, V. (1926). Variazioni e fluttuazioni del numero dindividui in specie animali conviventi. *Mem. Acad. Lincei Roma*, 2:31–113.

Waddell, J. N., Sander, L. M., and Doering, C. R. (2010). Demographic stochasticity versus spatial variation in the competition between fast and slow dispers. *Theor. Pop. Biol.*, 77:279–286.

Webb, S. D., Keeling, M. J., and Boots, M. (2007). Host-parasite interactions between the local and the mean-field: How and when does spatial population structure matter? *J. Theor. Biol.*, 249:140–152.

Wilkinson, R. R. and Sharkey, K. J. (2014). Message passing and moment closure for susceptible-infected-recovered epidemics on finite networks. *Phys. Rev. E*, 89:022808.

Willis, R. J. (2007). *The History of Allelopathy*. Springer.

Yates, C. A. (2014). Discrete and continuous models for tissue growth and shrinkage. *J. Theor. Biol.*, 350:37–48.

Yates, C. A., Baker, R. E., Erban, R., and Maini, P. K. (2012). Going from microscopic to macroscopic on nonuniform growing domains. *Phys. Rev. E*, 86:021921.

Young, S. H., Rozengurt, N., Sinnett-Smith, J., and Rozengurt, E. (2012). Rapid protein kinase D1 signaling promotes migration of intestinal epithelial cells. *Am. J. Physiol. Gastrointest. Liver Physiol.*, 303:G356–G366.

Young, W. R., Roberts, A. J., and Stuhne, G. (2001). Reproducing pair correlations and the clustering of organisms. *Nature*, 412:328–331.

Zheng, C. and Bennett, G. D. (2002). *Applied Contaminant Transport Modeling*. Wiley-Blackwell.