SPATIO–TEMPORAL MODELS FOR PLANT EPIDEMICS: ANALYTICAL AND SIMULATION STUDIES

Milena Maria Maule

Doctor of Philosophy
University of Edinburgh
2000

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

Acknowledgements

I would like to thank my supervisor, João Filipe, for his encouragement, support and constructive criticism throughout the course of this work.

I would also like to thank Gavin Gibson, Ben Hambly and Andreas Kyprianou for their valuable advice and many helpful discussions; the EPCC staff, and in particular Mario Antonioletti, for the technical support with the use of parallel computers; all BioSS students and staff; the Department of Mathematics and Statistics of the University of Edinburgh; and David Hill.

I gratefully acknowledge a research studentship from ICMS.

Abstract

In recent years the relation between spatial structure and temporal dynamics has become a central issue in population biology. We investigate the effects of different mechanisms of pathogen dispersal on the development of plant epidemics. Understanding how these mechanisms operate is useful in helping to predict and control epidemics in plant populations. Until recently spatio–temporal models often assumed that the transmission of disease occurs independently of the location of the individuals or is restricted to local contacts. We are interested in a wider range of dispersal processes which include various intermediate situations between these two extremes.

We formulate a stochastic spatio–temporal model for the spread of infectious diseases in plants. Studying the behaviour of a model which takes into account stochasticity and spatial extension usually involves intractable mathematics and requires the use of simulation. A challenging objective is to develop analytical methods for general application which provide predictions for the expected behaviour of the model.

The individual–based model comprises primary and secondary infection and recovery processes. Using stochastic simulation we study the expected behaviour and variability of the epidemic size, and characterise the disease patterns through spatial correlation. Both stationary and transient behaviour are analysed over the parameter space. Simulation is also used to test empirical extensions of non–spatial models which attempt to account for heterogeneous mixing of susceptibles and infecteds.

Analytical methods based on cluster approximations are commonly used for predicting the dynamics of stochastic models characterised by nearest neighbour (NN) interactions. On the other hand, for models with more general interactions, the rather simplistic and non–spatial Mean Field approximation has been extensively used. We propose an alternative general approach, built on individual–based ODEs and closure approximations, for predicting the behaviour of spatial models in which the individuals interact according to a generic function of their distance. The approximations, which take into account the development of correlations in the spatial distribution of the population, are tested against the simulation results showing excellent agreement in most of the parameter space. We also test the ability of cluster approximations to capture the effects of the anisotropic spread of the disease. To this end, we formulate a generalised NN model in which the dispersal of propagules depends on the direction of spread and use simulation to assess the performance of different approximations.

Table of Contents

Chapter 1

Introduction

1.1 Mathematical models for biological populations in space and time

A mathematical model of a biological or physical process is a representation of such a process designed to increase its appreciation and understanding. In general a mathematical model is meant to provide a quantitative description of the process under study so that predictions can be made about future events. In epidemiology, for example, mathematical models are constructed in an attempt to understand the way in which diseases are transmitted between individuals, to predict patterns of disease occurrence, to evaluate strategies for management and control of epidemics, and to provide insight into the design of efficient and cost–effective field experiments.

A variety of population dynamical models exist in the literature (see, for example, Renshaw, 1991). Probably the most important distinctions which should be made among these models are between deterministic and stochastic and between spatial and non–spatial (Mollison *et al.*, 1994).

Classical, deterministic and non-spatial population dynamical models are based on three fundamental assumptions (Bailey, 1975; Czárán, 1998), which are either explicit or implicit:

1. *abundance*: populations consist of a large number of individuals;

2. *uniformity*: all individuals of the same population are identical in every dynamically relevant respect;

3. *homogeneity*: each individual experiences the same environment around it. This assumption refers to perfect spatial mixing between individuals and is analogous to the assumption that in ideal gases identical particles

move fast and independently from each other and interact with probabilities proportional to the product of their densities.

In statistical physics, using these assumptions it is possible to describe successfully the macroscopic behaviour of gases given the microscopic properties of the particles of which they consist. In this context, assumptions 1–3 are not very restrictive, although they obviously fail to be realistic if applied to most biological systems. Nevertheless, by adapting the same approach to population dynamics problems, a number of phenomenological models were constructed which yielded deep theoretical insights into many aspects of population processes. In these models, all individuals belonging to the same species are assumed to be identical, forming a homogeneous mass. The basic objects are the categories or species of individuals in a population, and the relevant dynamical questions regard the abundances of each category or competing species. These models are also called "mass–action" (or mass–interaction) models because interactions occur at a rate directly proportional to the product of the number or density of the subpopulations involved (de Jong *et al.*, 1995). Some early examples of these models are the Malthusian exponential population growth, the Verhulst–Pearl model of logistic growth and the predator–prey interaction model by Lotka and Volterra (Renshaw, 1991).

Non–spatial deterministic models fail to describe the behaviour of biological populations because assumptions 1-3 are not as appropriate as they are for molecule populations.

The abundance assumption (1) is very often violated in that, even in very dense populations, the number of individuals is fewer by tens of orders of magnitude than the number of particles in a test tube. This simple example illustrates how, if at any time population numbers become small, a stochastic analysis becomes vital (Renshaw, 1991). Consider the Malthusian expression for epidemic development in a population where individuals develop independently of each other, reproduce at a constant rate λ and die at a constant rate μ. The number of individuals N at time t is then given by:

$$N(t) = N_0 e^{(\lambda-\mu)t}, \tag{1.1}$$

where N_0 is the initial size of the population. If births predominate ($\lambda > \mu$) the population will grow exponentially fast, whilst if deaths predominate ($\mu > \lambda$) the population will become extinct. For large N_0, and while $N(t)$ is large, this description is adequate; however, what happens if N_0 is small? Suppose $N_0 = 1$ and $\lambda = 2\mu$ (births are twice as likely as deaths), then the deterministic result

predicts exponential growth:

$$N(t) = e^{\mu t}. \tag{1.2}$$

Yet if the first event to occur is a death (probability: $\mu/(\mu + \lambda) = 1/3$), the population becomes extinct: thus the probability of ultimate extinction is at least $1/3$, which contradicts the deterministic prediction.

The uniformity assumption (2) is violated in every actual field case, almost without exception (Czárán, 1998): individual variation within biological populations is much greater than that of the particles in a general physical or chemical system. In principle, every individual of a community should be regarded as unique in terms of its combination of dynamically relevant genetical, physiological and environmental properties.

Unlike particles in a gas, individuals in a living environment move relatively slowly and do not mix homogeneously. Although in some relatively small communities some form of homogeneous mixing can be supposed, this assumption is completely unrealistic for widely spread populations. If, for example, we think about the spread of human diseases, it is easy to imagine that beyond the spatial scale of a household or a classroom the assumption that individuals mix homogeneously in space is not acceptable (Bailey, 1975). The homogeneity assumption (3) is even more evidently violated by plants (Gubbins & Gilligan, 1997 a, b) and benthic animals (Pascual & Levin, 1999), which almost do not move throughout their lifetime and for which interaction is very often limited to a neighbourhood of variable size.

The violation of the assumption of perfect homogeneous mixing shows the importance of including spatial components in models describing biological systems (Liu *et al.*, 1987). An illustrative example of the importance of spatial aspects can be found in one of the earliest representations of predator–prey behaviour (see Renshaw, 1991). Lotka (1925) and Volterra (1926) independently constructed a simple deterministic predator–prey model whose solutions are a family of closed curves in phase space, each curve corresponding to a different initial point. Motivated by this result, Gause (1934) made an empirical test of the model by rearing two species of protozoans (predators and prey) in an oat medium: in contrast to the theoretical prediction, he found that no matter how he altered the circumstances of his experiment, the predator always defeated the prey and the predicted cyclic behaviour did not occur. Simulations of the *stochastic* form of the Lotka–Volterra process confirm Gause's results, always leading to the extinction of one of the species. Once again the difference between stochastic and deterministic behaviours highlights the danger of making biological judgements based *solely* on

deterministic models. However, Huffaker (1958) questioned Gause's conclusions: he recognised that individuals rarely mix homogeneously and, by means of a large and complex *spatial* experiment, showed that the inability of the stochastic Lotka–Volterra model to generate sustained cyclic behaviour was only due to the absence of a spatial component. Moreover, computer simulations confirm that by including a spatial component in the stochastic Lotka–Volterra model, cyclic oscillations in the dynamics of the two populations may be restored.

Including spatial aspects in population dynamics enables the relaxation of at least the last of the three biologically unrealistic assumptions above, the homogeneity assumption (3). In spatial population models, the global spatial mixing of individuals is replaced by some spatially constraining mechanisms. There are in fact various ways to construct spatial mathematical models.

Durrett and Levin (1994 a) in their paper "The Importance of Being Discrete (and Spatial)" compare four different approaches to modelling the dynamics of spatially distributed systems. They consider non–spatial (*Mean Field*) approaches, where interaction between individuals is independent of their locations; *patch models*, where individuals are grouped in patches within which individuals interact homogeneously and between which there is migration independent of the distances between patches; *reaction–diffusion equations*, in which infinitesimal individuals are distributed in space and undergo purely local interactions; and *interacting particle systems*, in which individuals are discrete and space is treated explicitly.

Patch models, also known as pseudo–spatial models (Chesson, 1981), incorporate local (within patches) and global (between patches) dispersals, but interactions are typically independent of the distance between individuals or patches. Diffusion and reaction–diffusion equations (Okubo, 1980; Murray, 1989) are spatial but not discrete; the absence of discrete individuals, and hence of demographic stochasticity, makes them unable to predict fluctuations caused by the occurrence of random events in a finite population.

On the basis of these considerations, it is not surprising that in recent years one of the most successful and widely used approaches to spatio–temporal modelling has been that of interacting particle systems (IPSs) (Czárán, 1998). Much effort has been spent on demonstrating the importance of stochasticity, spatial extension and discreteness in models designed to represent real systems, where individuals do interact according to their relative locations in space and the processes occurring at the level of individuals are inherently stochastic (Jeger, 1989; Durrett & Levin, 1994 a; Shaw, 1994, 1995; Russell, 1996; Bolker & Pacala, 1997; Filipe & Gibson, 1998, 2000; Filipe *at al.*, 2000; Filipe & Maule, 2000 a, b).

The name "interacting particle systems" originates from the application of stochastic discrete–event models in physics (e.g. spin–glass models) to systems in which the interacting units were spatially fixed particles distributed on a regular lattice. The correspondence between lattice site and particle might be exact for many physical systems, but it is fuzzier in ecological and biological applications. For example, it may change with time, or the spatial scale of individuals and sites may not be the same. Five components are necessary to define the structure of an IPS model (Czárán, 1998):

1. a regular lattice of sites;
2. a finite set of possible cell states;
3. the size and the shape of the neighbourhood of each site, i.e. the set of sites affecting the state(s) of each site;
4. a transition rule defining how the state(s) of a site is changing;
5. an initial state configuration.

With simple local rules, IPSs are capable of generating very complex patterns at the lattice level, sometimes comparable to those produced by natural processes. Sometimes IPS models are also called Cellular Automata (CA) models. Strictly speaking however, CA are a special case of IPSs where the rules of local interaction are deterministic (Wolfram, 1986).

The problem with these types of models is that in general the only method for studying their behaviour is numerical simulation. Simulation of spatio–temporal stochastic models provides insights into the underlying mechanisms, possibly highlighting unforeseen features of a process and thereby suggesting further profitable lines of biological investigation. Computer simulations must be regarded as experimental mathematics, and simulation studies should be designed carefully and results analysed thoroughly (see Ripley, 1987). However, it is only possible to sample parts of the space of stochastic realisations of a simulated model. Moreover, this is computationally intensive. It is therefore desirable to develop mathematical and statistical approaches, providing complementary explanation and valuable guidance in how and what to investigate in simulations. Under certain assumptions, the dynamics of IPSs can be put in the form of difference or differential equations. Typically, these are not analytically solvable but can be simplified to approximated systems (Czárán, 1998; Filipe & Maule, 2000 b), leading to solutions which can then be compared to simulation of the full stochastic model.

1.2 Epidemic processes

Epidemiology is one of the most important areas of application of spatio–temporal modelling (Bailey, 1975; Mollison, 1995). The influence of infectious diseases on human history has repeatedly stimulated large–scale efforts to prevent, limit and eradicate the diffusion of epidemics. Mathematical and statistical models help to deepen our understanding of the processes involved in the spread of infectious diseases by providing insights into the mechanisms of transmission, the life–cycle of parasites and the effects of different control strategies. Modelling epidemics is not an easy task. Typically, epidemics involve processes at all scales, from global populations down to the individual level. The resulting dynamical systems are usually highly nonlinear and stochastic (Mollison, 1977).

Epidemiology has important applications not only to human populations but also to those of animals and plants. Understanding and controlling the spread of infectious diseases in natural and agricultural systems could have strong effects on economic, social and environmental aspects of our lives (Leonard & Fry, 1986). Modern agriculture faces the difficult challenge of finding a compromise between the necessity of increasing production to meet the growing demand for food and, at the same time, reducing the unsustainable impact such an increase has on the environment. The growth of the world population (at the present rate of increase, estimated to double in the next 50 years), the loss of productivity of crops due to reduction in the natural fertility of the soil, the emergence of pesticide–resistant pathogens, and the impact of modern agricultural methods on the environment are all problems that will have to be tackled. This will require a lot of progress in our understanding of a large number of different processes, among which are the spread of infectious diseases in plant populations and the design and evaluation of possible strategies for control.

A disease is said to be infectious (Bailey, 1975) if transmission occurs, at some stage in the life–cycle of the pathogenic organism, from an infected host to an uninfected susceptible, with or without the mediation of an intermediate vector. The biological process underlying the spread of many infectious diseases consists of a regular cycle of events. Consider an individual which has been *exposed* to infection, for example through physical contact with infectious individuals or material. The individual may be resistant to the disease to some extent because of its own biological defences (which could have been acquired in response to previous exposure) and not become ill. If it is not immune to the disease, the invading parasites develop and the individual undergoes a *latent* period in which it is contracting the disease but does not discharge any infectious material. After this

latent period, the individual is capable of communicating infectious organisms to other susceptibles. At this stage the infected individual is *infective*, and the time during which it communicates infecting organisms is called the *infectious* period. At some stage in the individual's history of infection, recognizable symptoms of the disease appear. The time from the receipt of the infection and the occurrence of the symptoms is called the *incubation* period, and the time from the observation of symptoms in one case to the observation of symptoms in a second case is called the *serial interval.* More complicated cycles of infection processes can occur, for example in the presence of carriers (apparently healthy individuals capable of transmitting infection to others) or vectors (when the infective agent spends part of its life–cycle inhabiting an intermediate host).

A common way to represent the infection process described above is the construction of a compartment model where each compartment (or class) represents a category of individuals (Mollison, 1995). In the most generic model there are four classes of individuals:

- **S**: Susceptible. Individuals capable of contracting the disease.
- **E**: Exposed. Individuals which are infected but not infectious.
- **I**: Infective. Individuals capable of transmitting the disease.
- **R**: Recovered (or Removed). Individuals immune from the disease.

Fig. 1.1 shows a schematic visualisation of the model. S, E, I, R represent the

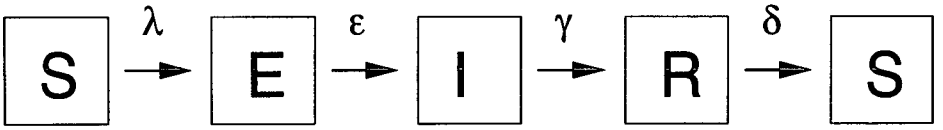


Figure 1.1: *Visualisation of the SEIRS compartment model.*

fractions of the population (or numbers of individuals) belonging to each of the categories listed above and satisfy $S+E+I+R=1$ (or $S+E+I+R=N$, with N the total number of individuals). Susceptible individuals become exposed at rate λ, then infective at rate ϵ, then recovered with temporary immunity at rate γ, and then susceptible again, when immunity is lost, at rate δ.

Once a population has been divided into various compartments with different properties, deterministic and stochastic versions of an epidemic model can be formulated. In the context of deterministic models, given knowledge of the

relations between the various sub–populations, a system of equations can be set up to model the dynamics of the whole population. Using discrete time intervals, the model would take the form of a set of difference equations, whilst if instantaneous rates were used, the model would form a set of differential equations. A stochastic model, instead, treats the population variables as being random variables. In this case, differential equations could be used to model the evolution of the probability densities of the system variables (Bailey, 1975). Systems of differential equations corresponding to the stochastic and the deterministic versions of nonlinear models are therefore different and usually lead to different conclusions (Mollison, 1977; Gibson *et al.*, 1999). In order to illustrate this point we may consider the following simple example. A population consists of N individuals. At any given time t there are $I(t)$ infected individuals and $S(t)$ susceptibles, and $N = I(t) + S(t)$. Infecteds transmit the disease to susceptibles at a constant rate λ. Hence, the occurrence of new infections is governed by the probabilistic rule

$$\mathbf{P}[I(t + dt) = I(t) + 1] = \lambda\ I(t)S(t)\ dt. \tag{1.3}$$

Using $\langle \cdots \rangle$ to denote expectation, we can write the equation for the time evolution of the expectation of $I(t)$ in the previous equation as

$$\frac{d\langle I(t)\rangle}{dt} = \lambda\ N\ \langle I(t)\rangle - \lambda\ \langle I(t)^2\rangle, \tag{1.4}$$

where we used $S(t) = N - I(t)$. The dynamic equation corresponding to the deterministic version of the model is

$$\frac{dI(t)}{dt} = \lambda\ N\ I(t) - \lambda\ I(t)^2. \tag{1.5}$$

Since $\langle I^2\rangle \neq \langle I\rangle^2$, the dynamics of equations (1.4) and (1.5) are distinct, so that the deterministic version of the model cannot be interpreted to describe the expectation of disease levels.

A number of special cases of the general SEIRS compartment model illustrated in Fig. 1.1 should be considered (see, for example, Liu *et al.*, 1987):

1. if there is no latent period (no exposed class E), we obtain the SIRS model where infected individuals are infective (this corresponds to the limiting case where the expectation of the latent period $1/\epsilon$ tends to 0 in the stochastic model);

2. if there is no immunity (no recovered class R), we obtain the SEIS model (limiting case where the expectation of the period of immunity $1/\delta \to 0$ in the stochastic model);

3. if immunity is permanent, then the average period of immunity $1/\delta$ is infinite, so the SEIRS model is reduced to the SEIR model and the SIRS to the SIR;

4. if there is no latent period and no immunity upon recovery, we obtain the SIS model (limiting case of the SIRS model as $1/\delta \to 0$, or limiting case of the SEIS model as $1/\epsilon \to 0$);

5. finally, if there is no latent period and no recovery, we obtain the SI model (limiting case of the SIS model when γ is infinite).

1.3 Spatio–temporal models for the spread and control of infectious diseases in plant populations

The objective of this PhD research project is the investigation of spatio–temporal models for the propagation and control of plant disease epidemics.

We consider spatio–temporal stochastic models which fall into the class of IPS models. Our primary intention is to gain understanding of the spatial properties of the spread of infectious diseases. Traditionally, many epidemiological models have considered interactions between individuals which are either independent of their relative locations (Mean Field models) or restricted to local contacts, for which the probability of disease transmission decays exponentially with distance or only occurs between nearest neighbours (Harris, 1974; Levin & Durrett, 1997). It is likely that the behaviour of real dispersal processes lies between these two extremes. In fact, while infection may be a fairly localised process well described by rapidly decaying interactions, for many diseases the dispersal of propagules such as airborne spores or viruses carried by vectors can occur over large distances (Minogue, 1989; Gibson, 1997). There is evidence to support the use of long–tailed *distribution functions* (dispersal functions or contact distributions (Mollison, 1977)) such as power–laws for describing the dispersal of pathogens (Fitt & McCartney, 1986; Shaw, 1995).

Motivated by a paper by Gibson (1997), where stochastic models were fitted to observations of aphid–borne virus (*citrus tristeza virus*) epidemics in orchards, we construct simple SI and SIS models in which the contact distributions decay with distance as power–laws and cover a spectrum of possible dispersal scenarios. In particular, we concentrate on cases of long, medium and short distance dispersal and investigate the corresponding differences in the development and pattern of disease (Filipe *et al.*, 2000; Filipe & Maule, 2000 a).

The relevance of these models is not restricted to plant epidemiology. One of the reasons for considering general and biologically simple models resides in the possibility of extending the findings of this study to other areas of interest, such as the dispersal of plant genetic material (Mollison, 1977; Lewis, 1997).

Despite their simplicity, these models incorporate significantly realistic features, such as spatial components, discreteness of individuals and stochasticity. Such characteristics make them appropriate for describing processes at the small spatial and temporal scales of many agricultural crops and natural systems. However, the introduction of spatial components leads to intractable mathematics whose solution requires the use of approximations.

We begin by investigating the model behaviour through computer simulations. Here we look at the number and spatial distribution of infectives and how they vary with the model parameters, system size and boundary conditions.

The stochastic nature of events and the type of dispersal processes, which allow each individual to interact with every other, make simulation of the model computationally very intensive. It is therefore desirable to develop simpler analytical descriptions for examining the behaviour of the model. Semi–spatial approaches are in general limited to lattice models with very restricted interaction neighbourhoods (e.g. Sato *et al.*, 1994; Filipe & Gibson, 1998, 2000). On the other hand, Mean Field approaches have been extensively used for representing all other forms of interaction. We develop analytical methods for the dynamics of models with general interactions and test the accuracy of their predictions against simulation. The method is based on a set of deterministic differential equations which describe the essential dynamics of the stochastic model, and on closure approximations. Moment-closure approximations, recently proposed by Bolker and Pacala (1997) and Bolker (1999) in the context of SI, SIS and SIR models, are found to be inadequate when interaction rates are not small compared to birth-death rates. More robust closure schemes are investigated and found to give predictions in very good qualitative and quantitative agreement with the simulation of the full stochastic model over most of the parameter space (Filipe & Maule, 2000 b).

The potential of this analytical approach goes beyond the mere purpose of predicting the behaviour of the stochastic models so far considered. The same method can in principle be applied to any lattice based stochastic model with spatial interactions of any type. Moreover, provided that the set of relative distances between individuals (or particles) is known, the model need not be based on a regular lattice. In addition, the mathematical clarity of the approach makes it very flexible for accommodating modifications of the model needed to represent

other systems.

Motivated further by interest in the spatial properties of dispersal processes (of infectious propagules or other vegetation material, such as pollen or genes), we consider a generalisation of the nearest neighbour (NN) model which includes spatial anisotropy in the interaction between individuals. Here the dispersal of propagules from infectives to neighbouring sites depends on the direction of spread. Anisotropy can occur in a variety of real situations; examples include the effect of wind on the dispersal of airborne spores (Zawolek, 1993), the spread of soil–borne diseases in fields whose morphology affects the flow of water (Hickman, 1940), and the effect of rectangular rather than square plots where distances between columns and rows of plants are different, favouring interactions in one particular direction (Aylor & Ferrandino, 1989). The behaviour of the anisotropic NN model is investigated using computer simulation and various *cluster approximations* aimed at obtaining analytical predictions (Bethe, 1935; ben Avraham & Köhler, 1992; Filipe & Gibson, 1998, 2000). The ability of the different approximations to capture anisotropic behaviour is assessed.

The final model which we consider is also aimed at accounting for spatial components of epidemics, but it does so in implicit rather than explicit form. Standard deterministic epidemiological models rely on the mass–action principle which ignores the spatial dimension. There the incidence rate of infection is assumed to be bilinear in the fraction of infectives and susceptibles of the population (Bailey, 1975). This assumption implies that there is homogeneous mixing between all individuals in the population (see assumption (3) at the beginning of this chapter). One possible way to relax this assumption without explicitly including space is to modify the incidence rate so that the fractions of infectives and susceptibles appear raised to a power different from 1, thereby introducing additional nonlinearities (Liu *et al.*, 1987; Gubbins & Gilligan, 1997 a, b). This model is found to improve the fit to experimental observations. However, it is not clear whether the goodness of fit is a mere consequence of the extra parameters or whether there is a deeper relation between these parameters and spatial heterogeneity. We attempt to address this question in the context of our simple but spatially explicit model, and derive a formal relation between the parameters in the nonlinear incidence rate model and the autocorrelation functions of the spatial model. This relation is then tested using simulation.

1.4 Overview

This chapter has introduced the topics of the work of this thesis and the motivations that led us to them.

In chapter 2 the stochastic spatial model for epidemic spread and control is defined. Some definitions and terminology are introduced. A set of ordinary differential deterministic equations describing the essential dynamics of the model are derived. Although these ODEs are given specifically for the model under consideration, they have general application and form the basis for deriving analytical approaches for different types of IPS models.

Chapter 3 investigates the behaviour of the stochastic model using computer simulation. Both SI and SIS models are considered and the parameter space is explored thoroughly. In particular, cases of short, medium and long range dispersal are considered. The variation of observables of interest with different parameter values, system sizes, initial conditions and boundaries is analysed.

Chapter 4 reviews analytical methods for models with NN interactions. Cluster approximations of different orders are presented and their predictions compared to simulation.

In chapter 5 the NN model is extended to include spatial anisotropy in disease spread. The effects of anisotropy on epidemic development are studied using computer simulation. In addition, the approximation techniques for the isotropic NN model (chapter 4) are generalised to the current model, and their ability in capturing the new features of behaviour is assessed.

In chapter 6 analytical methods are developed in order to predict the expected behaviour of stochastic population models with general spatial interactions. The performance of various approximations is evaluated through comparison with the simulated data studied in chapter 3. In particular, the moment–closure approximation of Bolker & Pacala (1997, 1999) is shown to be less appropriate than others proposed by us.

Chapter 7 considers an empirical model for heterogeneous mixing of infectives and susceptibles proposed by other authors (Liu *et al.*, 1987; Gubbins & Gilligan, 1997 a, b). The model is tested in the context of the model from chapter 2 and the relation between the empirical parameters and spatial correlations is studied.

Chapter 8 summarises the main findings and discusses possible further developments of this work.

Chapter 2

A model for the spread and control of infectious diseases in plants

2.1 Introduction

This chapter introduces a stochastic spatio–temporal model for the propagation and control of plant disease epidemics. We discuss and interpret the salient characteristics of the model and explain the meaning of the parameters which determine its behaviour. Finally, we derive a system of ordinary differential equations which describes the model dynamics.

The epidemic model considered is very simple: susceptible individuals can acquire the disease from external sources of infection or through transmission from infected individuals within the population; at the same time, infecteds are replaced by healthy individuals at a constant rate (Filipe *at al.*, 2000; Filipe & Maule, 2000 a). Although simple, the model comprises two important and realistic features, namely stochasticity and spatial extension. Stochasticity is essential when dealing with biological systems at small geographical or short temporal scales, or with populations consisting of a small number of individuals. The importance of spatial extension becomes evident if we think that individuals interact according to their relative location in space. In recent years a great effort has been made to emphasise the importance of both these aspects (Mollison, 1977; Durrett & Levin, 1994 a, b; Shaw, 1995; Gibson, 1997; Gibson *et al.*, 1999).

One of the main objectives of this thesis is to study the relation between the mechanisms of pathogen dispersal and the disease patterns which they generate. The dispersal of infectious propagules in plant populations determines the spatial distribution of disease. This in turn affects the development of epidemics and their persistence. Traditionally, epidemiological models have considered two contrast-

ing situations: interactions independent of the location of hosts, and interactions restricted to contact between neighbours (Harris, 1974; Durrett & Levin, 1994 b). Since it is plausible that real dispersal processes lie between these two extremes, our model comprises a whole range of possible interactions.

In order to characterise the model behaviour we define a number of observables of interest, such as epidemic size and spatial correlations, and describe the sampling methods we shall use to estimate them by means of stochastic simulation.

An alternative and complementary approach to lengthy simulation studies is the development of analytical methods for exploring and predicting the behaviour of the model. We derive a set of ordinary differential equations (ODEs) describing the dynamics of our epidemic model. Approximation techniques for closing and solving such a system of equations will be the subject of chapter 6. From the analysis of these individual–based dynamic equations some characteristic features of the model behaviour can be predicted. The formulation of an analytical method capable of providing approximate solutions has much wider application than the example of plant epidemics considered here. Any individual–based population model characterised by general spatial interactions can be studied using a similar approach in order to obtain evolution equations for expectations, whose solution might require the use of approximation techniques.

2.2 Spatio–temporal stochastic model

We consider a population of plants distributed on a square lattice of linear size L. Each lattice site is occupied by one individual that is either susceptible or infected. The state of the individual located at site $\mathbf{x}$ is specified by a binary variable $S_{\mathbf{x}}(t)$ which depends on time t and is equal to 0 if the individual is susceptible and 1 if it is infected:

$$S_{\mathbf{x}}(t) = \begin{cases} 0 & \text{Susceptible} \\ 1 & \text{Infected.} \end{cases}$$

Two types of lattice systems are considered: unbounded, with toroidal periodic boundary conditions (PBC), and bounded, with fixed boundary conditions (FBC) on a square with no infecteds outside the boundary.

The population dynamics are defined as follows. At a given time t, a susceptible $\mathbf{x}$ can acquire the disease through background (or *primary*) infection from external sources at a constant rate J_1, or through *secondary* transmission from an infected $\mathbf{y}$ inside the population at a rate $J_2 f(|\mathbf{x}-\mathbf{y}|)$, where $f(|\mathbf{x}-\mathbf{y}|)$ is a monotonically decreasing function of the distance $|\mathbf{x}-\mathbf{y}|$. The probability of

infection of $\mathbf{x}$ in a small time interval dt is

$$\mathbf{P}\left[S_{\mathbf{x}}(t+dt)=1|S_{\mathbf{x}}(t)=0\right]=\left[J_1+J_2\sum_{\mathbf{y}\neq\mathbf{x}}f(|\mathbf{x}-\mathbf{y}|)S_{\mathbf{y}}(t)\right]dt. \tag{2.1}$$

In addition, each infected recovers (or is replaced by a susceptible) with probability

$$\mathbf{P}\left[S_{\mathbf{x}}(t+dt)=0|S_{\mathbf{x}}(t)=1\right]=R\,dt, \tag{2.2}$$

thus $1/R$ is the mean infectious time. In this model, since there is no latent period, infected also means infectious.

2.2.1 Parameters of the model

Recovery is a common process in models with local contacts (Harris, 1974). It can be regarded as a strategy for control through the replacement of diseased individuals with healthy ones (such as in Filipe & Gibson, 1998; Filipe *et al.*, 2000). It can also represent the death of infecteds quickly followed by the birth of healthy offspring. A similar process was considered, for example, by Shaw (1995). We studied the model with and without recovery. When recovery is present ($R > 0$), the model falls into the class of SIS (Susceptible–Infected–Susceptible) models; without recovery ($R = 0$), the model belongs to the class of SI (Susceptible–Infected) models.

Since in most cases plant populations are not completely isolated, it is plausible to imagine that some infectious propagules are transported into the system from outside. A model with both primary and secondary infection was found to provide a good fit for observations of epidemics of the citrus tristeza virus in orchards (Gibson, 1997). Background infection is often present, even if at low levels (small J_1). This has a randomisation effect on the pattern of disease, because individuals may become infected independently of their spatial location.

J_2 measures the strength of the interaction between infecteds and susceptibles. It can be thought of as the rate of propagule production by each infected in the population.

2.2.2 Contact distribution

The function $f(|\mathbf{x}-\mathbf{y}|)$ describes the interactions within the population and has been named the *contact distribution* (Mollison, 1977) or the *dispersal function* (Minogue, 1986). It represents the probability distribution for the displacement between the source and the point of deposition of a propagule. Many spatial

population models have often assumed that interactions result from local processes, such as contacts between neighbours (Harris, 1974; Durrett & Levin, 1994 b), exponentially decaying dispersal functions (Minogue & Fry, 1983) or diffusion (Zadocks & Van den Bosch, 1984). On large enough scales, these models exhibit qualitatively similar behaviour, such as well–defined spreading clusters (Mollison, 1977). In many cases infection is actually a localised process, well–described by such interactions. Nevertheless, for many infectious diseases, propagules can travel over large distances to produce new infections which act as secondary sources of spread. This is the case of air–borne spores or insect–borne viruses, which can be transported over unlimited distances (Fitt & McCartney, 1986). Both power–law and exponential contact distributions can describe a wide range of dispersal processes, from nearest neighbour (NN) to uniform, non–spatial interactions. Experimental counts of fungal spores have been successfully fitted with power–law dispersal functions (Mundt & Leonard, 1985; Fitt & McCartney, 1986), although the insufficient number of counts at large distances often implies that exponentials can also explain the data (Minogue, 1989). The simulation of models reveals that the patterns of disease produced by the two types of dispersals are very different (Mollison, 1977; Minogue, 1989; Shaw, 1995). Power–law contact distributions generate patchy patterns of diseases with no well–defined epidemic fronts, which are more complex and realistic than those produced by exponentials (Minogue, 1986; Shaw, 1995). More direct experimental evidence for the use of power–laws comes from the statistical analyses of observations of viral disease epidemics in citrus trees by Gibson (1997) and Gibson & Austin (1996).

Motivated by these considerations, we assume that the contact distribution decays with the distance $r = |\mathbf{r}|$ as a power–law: $f(r) \propto r^{-a}$. The exponent a determines the range of the dispersal and reflects the nature of the mechanism of dispersal. When a is small, dispersal is long–ranged. In particular, for $a = 0$ dispersal is independent of the spatial distribution of individuals and propagules are deposited randomly. The range of dispersal decreases as a increases. When a is large, dispersal is short–ranged, and in the limit in which $a \to \infty$ interaction is restricted to nearest neighbours. In this limit the model corresponds to the Contact Process introduced by Harris (1974) in probability theory to describe interactions on a lattice. In the literature relevant to plant disease epidemiology, typical values for the exponent a vary from 1.1 to 6 (Minogue, 1986).

To ensure the conservation of the number of propagules emitted from a source and to obtain desirable asymptotic properties, the contact distribution is nor-

malised over all possible depositions on a dispersal domain D:

$$f(|\mathbf{r}|) = \frac{|\mathbf{r}|^{-a}}{Z(a,D)}, \quad \mathbf{r} \in D, \tag{2.3}$$

where

$$Z(a,D) = \sum_{\mathbf{s}\in D} |\mathbf{s}|^{-a}, \tag{2.4}$$

so that $\sum_{\mathbf{r}\in D} f(|\mathbf{r}|) = 1$. We assume that each infected produces the same amount of infectious spores and that the dispersal domain does not depend on the location of the source. D must then be the same for periodic and bounded systems. Spores can be deposited on a susceptible or infected site within the system, or outside the system if there is a boundary. In practice we consider dispersal up to a truncation distance r_{max} from the source. If r_{max} is large, this truncation is justified because the fraction of propagules deposited beyond a certain distance from the source can be considered negligible. We define the dispersal domain D as a square centred at the source with side $2L$ and area four times larger than the system, such that D is the set of vectors $\mathbf{r}$ with $|\mathbf{r}| > 0$ and Cartesian components less than L. With this definition, interaction is possible between all sites with either boundary conditions. When $a \leq 2$, Z diverges as the size of the dispersal domain tends to ∞, and truncation is necessary. For $a > 2$, instead, Z has a finite limit and, for the parameter values we considered, there is very little difference between this limit and the truncated Z.

The contact distribution defined in (2.3) does not have a parameter setting a dispersal length–scale. A range of dispersal can be defined by the median m of the distribution (Minogue, 1989; Shaw, 1995). An estimate of m can be obtained by approximating the lattice system with a spatially continuous system and changing the sums into integrals. With r_0 indicating the lattice space unit, m is the solution to the equation

$$F(m) \equiv \frac{2\pi}{Z} \int_{r_0}^{m} r^{1-a} dr = \frac{1}{2}, \tag{2.5}$$

which is

$$\frac{m}{r_0} = \begin{cases} \left[\dfrac{2}{1+(r_0/L)^{a-2}}\right]^{\frac{1}{a-2}} & a \neq 2 \\ \sqrt{L/r_0} & a = 2. \end{cases} \tag{2.6}$$

The median has the advantage of being finite when the system size L tends to ∞ for $a > 2$, whilst the mean requires $a > 3$ to be finite. Fig. 2.1 shows the behaviour of the median m for a lattice of size $L = 200$ and lattice unit $r_0 = 1$ as a varies from 0 to 20.

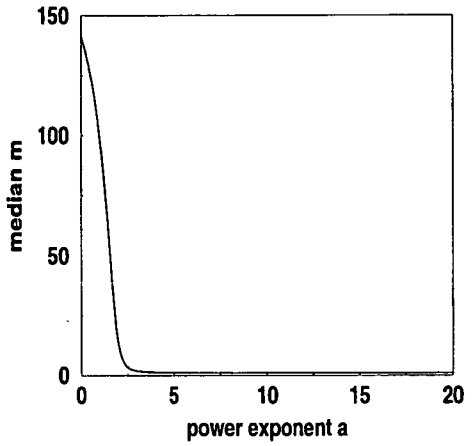


Figure 2.1: *Median m vs the exponent a of the contact distribution for a lattice of size $L = 200$ and lattice unit $r_0 = 1$.*

2.3 Structure of the lattice

Lattices are often used in mathematical models to reduce the computational difficulties of problems formulated in continuous space. Lattice models are straightforward to simulate and, by definition, prevent multiple occupancy of the same spatial location by several individuals. Moreover, in agricultural systems, plants are typically sown along sets of rows which are well described by a lattice. Instead, natural systems usually have less regular spatial distributions. Nevertheless, the overall effect of lattices on the behaviour of the model should not be important on the scale of large systems.

We measured $G(r)$, the spatial–average number of sites at distance r from a given site, both in lattices with and without boundaries. The situation is different for the two systems because in the lattice with FBC the distribution of sites around a given site $\mathbf{x}$, $G_{\mathbf{x}}(r)$, depends on the location of $\mathbf{x}$, whilst with PBC each site is surrounded by the same lattice environment and $G_{\mathbf{x}}(r)$ coincides with its spatial–average $G(r)$. For example, on a square lattice of linear size $L = 200$ and PBC, each of the $40,000$ sites is surrounded by 4 neighbours at distance 1 lattice unit, hence $G(1) = 4$; on a lattice with FBC, $198^2 = 39,204$ sites are surrounded by 4 neighbours, $198 \times 4 = 792$ sites by 3 neighbours and 4 sites (in

the corners) by 2 neighbours, hence $G(1) = 3.98$. Fig. 2.2 shows $G(r)$ for lattices of linear size $L = 200$ with each type of boundary condition.

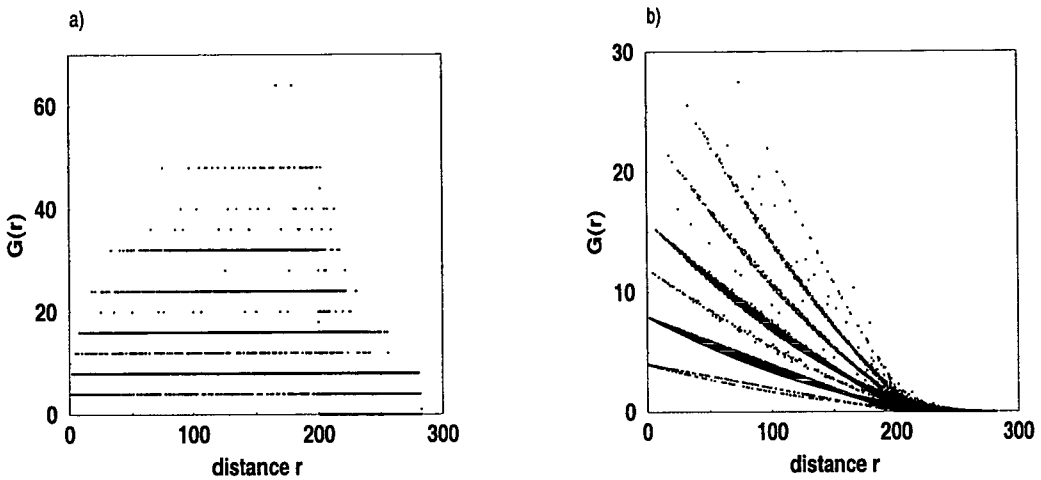


Figure 2.2: $G(r)$, *the distribution of sites at distance* r *from a given site averaged over all lattice sites. a) PBC; b) FBC.*

The meaning of $G(r)$ is better understood if we plot its histogram

$$H(r) = \sum_{s \in [r-\Delta/2, r+\Delta/2[} G(s), \tag{2.7}$$

where r varies between the lattice unit ($r = 1$) and the maximum distance between two sites in the system ($r = \sqrt{2}L$). Histograms with windows of width 1 of the average radial distribution for both types of lattices are shown in Fig. 2.3. The histogram for the unbounded system follows closely the line $2\pi r$, the distribution corresponding to a continuous system.

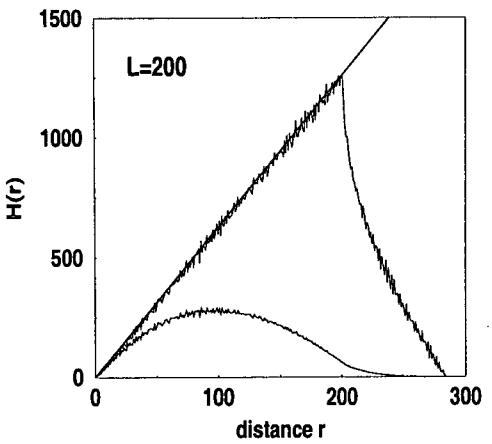


Figure 2.3: *Histograms of the radial distribution of sites for lattices with PBC (top) and FBC (bottom). The straight line with slope* $2\pi r$ *is the continuous radial distribution.*

2.4 Statistical analyses of model behaviour

The model comprising both disease transmission and recovery exhibits two different regimes of behaviour. Initially, the fraction of infecteds (or epidemic size) is subject to random fluctuations and tends to increase. This stage is called the *transient regime*. In finite systems the probability of extinction is non–zero (Liggett, 1985): provided that we wait long enough, a sufficiently large downward fluctuation in the epidemic size will occur which drives the epidemic to extinction. However, if a long time has to elapse before the probability of extinction becomes non–negligible, the system may reach a long–term *quasi–stationary regime*. In this stage, the size of the epidemic fluctuates around a quasi–equilibrium value which is close to the corresponding equilibrium epidemic size of an infinite system. We shall refer to this phase either as the stationary, or the equilibrium, or the long–term, or the asymptotic regime (i.e. we drop the prefix "quasi"), keeping in mind that true stationarity can only be reached in infinite systems.

When recovery is not present, the epidemic grows until the whole population is infected, in which case there is no stationary regime.

Transient and stationary regimes have distinct spatial and temporal characteristics (Filipe *et al.*, 2000; Filipe & Maule, 2000 a). In particular, transient regimes are difficult to characterise because they can be highly stochastic and are strongly affected by initial conditions, i.e. the initial number and spatial distribution of infecteds in the system. However, in practice, transient regimes are very important. In fact, real systems may never reach stationarity during the relevant periods of observation: for example, seasonal crops may be harvested before an epidemic reaches its equilibrium level. On the other hand, the non–degenerate stationary regime of the model is a unique, ergodic state, independent of the initial state of the system, whose existence has been rigorously proven for the Contact Process (Harris, 1974; Liggett, 1999).

We investigate both stationary and transient regimes by means of computer simulation. Each stochastic realisation of the process consists of a simulated time series $\{t_i, \mathbf{S}(t_i)\}$, with t_i $(i = 1, \dots)$ the instants in continuous time when an event occurs (i.e. a change in the state of one individual), and $\mathbf{S}(t_i) = \{S_{\mathbf{x}}(t_i) | \mathbf{x} \in$ system$\}$ the population of random variables representing the states of all $N = L^2$ individuals in the system during the waiting time $\Delta t_i = t_{i+1} - t_i$. At any given time t, two levels of statistics are involved: spatial statistics, over the population of N individuals, and realisation statistics, over a random sample of n time series $\{t_i, \mathbf{S}(t_i)\}_s$ $(s = 1, \dots, n)$, drawn from the appropriate space of stochastic realisations. Samples of realisations are drawn conditional on epidemic

survival.

We describe first the realisation statistics. Consider a random variable $A_s(t)$, where the index s labels one of the n stochastic realisations ($s = 1, \ldots, n$), and which depends on some or all the variables $\mathbf{S}_s(t)$. In the transient regime, its expectation, $\langle A(t) \rangle$, is estimated by the average over the sample of realisations:

$$\langle A(t) \rangle \approx \langle A(t) \rangle_n = \frac{1}{n} \sum_{s=1}^{n} A_s(t), \tag{2.8}$$

where

$$\langle A(t) \rangle = \lim_{n \to \infty} \langle A(t) \rangle_n. \tag{2.9}$$

The sample variance is given by

$$\sigma^2[A(t)]_n = \langle A(t)^2 \rangle_n - \langle A(t) \rangle_n^2. \tag{2.10}$$

Ergodicity implies that, in the stationary regime, conditional on epidemic persistence, it is equivalent to sample either over time in a single random realisation or over many realisations. Hence, denoting by A the long–term stationary behaviour of $A(t)$, we estimate its expectation over realisations as

$$\langle A \rangle \approx \langle A \rangle_n = \frac{1}{n} \sum_{s=1}^{n} \left[\frac{1}{t_{obs}} \int_{t_{max} - t_{obs}}^{t_{max}} A_s(t') dt' \right], \tag{2.11}$$

where t_{obs} is an observation time during the stationary regime and t_{max} is the maximum duration of each realisation. Since $A_s(t)$ is a step function, the integral equals the sum of $A_s(t_i)$ weighted by $\Delta t_i / t_{obs}$, the relative duration of each state. The number of stochastic realisations n needed to estimate the average over the realisation space is much larger in the transient regime than in the stationary regime.

For each stochastic realisation at given time t, we considered two spatial statistics:

1. the fraction of infecteds (or epidemic size)

$$\rho_I(t) = \frac{1}{N} \sum_{\mathbf{x}} S_{\mathbf{x}}(t), \tag{2.12}$$

which is the ratio of the number of infecteds to the total number of individuals;

of infecteds. Since in our model the disease is transmitted from infecteds to susceptibles, each individual can influence the state of the others and, as a result, correlations develop. This effect clearly depends on the distance between individuals and tends to zero when the distance becomes very large, i.e. the spatial range of correlations is finite. However, the range of correlations may become quite large and even tend to infinity in the proximity of a persistence threshold.

Correlations should not be confused with interactions. Interactions always produce correlations, but correlations can also be produced by other means, such as the initial preparation of the system. Short–range interactions may give rise to long–range correlations as a result of co–operative effects involving chains of consecutive nearest neighbours.

A spatial autocorrelation function can be defined in different ways. We first define the expectation over realisations of the correlation between two sites $\mathbf{x}$ and $\mathbf{y}$ in the lattice at a given time t,

$$C_{\mathbf{xy}}(t) = \frac{\langle [S_{\mathbf{x}}(t) - \langle S_{\mathbf{x}}(t)\rangle]\,[S_{\mathbf{y}}(t) - \langle S_{\mathbf{y}}(t)\rangle]\rangle}{\sigma[S_{\mathbf{x}}(t)]\;\sigma[S_{\mathbf{y}}(t)]}, \tag{2.17}$$

where $\sigma[S_{\mathbf{x}}(t)] = \sqrt{\langle [S_{\mathbf{x}}(t) - \langle S_{\mathbf{x}}(t)\rangle]^2\rangle}$ is the standard deviation of the variable $S_{\mathbf{x}}(t)$.

Correlations between specific individuals are not very informative. The spatial average of this quantity, however, tells us how correlations decay with the distance. To obtain this average we consider all pairs of sites $\mathbf{x}$ and $\mathbf{y}$ at a given distance r,

$$C(r,t) = \frac{1}{N\,G(r)} \sum_{\mathbf{x}} \sum_{\mathbf{y}:|\mathbf{x}-\mathbf{y}|=r} C_{\mathbf{xy}}(t). \tag{2.18}$$

We shall use this estimator to evaluate correlations in bounded, spatially heterogeneous lattices.

In the special case of spatially–homogeneous, unbounded lattices, (2.18) can be rewritten in a simpler form. Translational invariance ensures that $\langle S_{\mathbf{x}}(t)\rangle = \langle S_{\mathbf{y}}(t)\rangle = P_I(t)\ \forall\ \mathbf{x},\mathbf{y} \in$ system, and that $\langle S_{\mathbf{x}}(t) S_{\mathbf{y}}(t)\rangle = P_{II}(r,t)$ only depends on the distance $r = |\mathbf{x}-\mathbf{y}|$ between the two sites. Moreover, recalling that $S_{\mathbf{x}}(t)$ is a binary variable which takes values 0 or 1, i.e. $S_{\mathbf{x}}(t)^2 = S_{\mathbf{x}}(t)$, (2.18) can be written as

$$C(r,t) = \frac{P_{II}(r,t) - P_I(t)^2}{P_I(t) - P_I(t)^2}. \tag{2.19}$$

The denominator in (2.19) can be thought of as the limit of the numerator as $r \to 0$. From (2.19) it is easy to see that the spatial autocorrelation function

quantifies how much a pattern of disease deviates from a random pattern. The first term in the numerator represents the probability of finding two infecteds at distance r at time t. For a random distribution, this probability is given by the square of the probability of finding one infected at time t, i.e. $P_I(t)^2$. The numerator of (2.19) is simply the subtraction of these two terms.

There are other estimators of the expectation of autocorrelation functions. In particular, we compared (2.19) for a stationary system with $\langle(\rho_{II}(r) - \rho_I^2)/(\rho_I - \rho_I^2)\rangle$ but found very little difference between them.

Typically, for this model, the empirical measurements of autocorrelations exhibit an initial smooth decay followed by a noisy tail symmetric around the r-axis. The main source of noise is the spatial average over pairs at a given distance, $P_{II}(r,t)$ in (2.19), and the double summation in (2.18). Increasing the linear size of the system (for example from $L = 100$ to 200) reduces noise substantially. Noise also originates from the irregular radial distributions $G(r)$ and $G_{II}(r)$, a lattice effect which can be reduced using histograms (see Fig. 2.3). We therefore smoothed (2.19) by replacing $G_{II}(r)$ and $G(r)$ in $P_{II}(r) = \langle G_{II}(r)/G(r)\rangle$ with histograms (with windows of width 1 or 2), as in (2.7).

2.5 Analytical approach

Spatio–temporal stochastic models usually lead to intractable mathematics where even approximate solutions can be difficult to obtain. The usual answer to this problem is to investigate the model behaviour using computer simulation. However, the inherent stochasticity of the model introduced in this chapter and the long–range nature of the interactions make this task particularly difficult and computationally intensive. Moreover, although simulation models are very flexible, they are also very difficult to analyse fully, making it hard to generalise results obtained for a particular model.

We wish to derive dynamic equations which describe the behaviour of the stochastic model introduced in the previous sections. The importance of these equations goes beyond the sole purpose of devising an alternative description of our epidemic model. They are in fact the basis for the development of analytical methods capable of dealing with spatial and stochastic aspects of general population models where individuals interact according to a generic function of their distance. Approximation techniques are used to solve these equations and obtain deterministic predictions for the model behaviour (see chapter 6 and Filipe & Maule, 2000 b). These predictions are then compared to simulation results. They can be used to test hypotheses or to fit models to observations.

In the following sections, dynamic equations describing the evolution of $P_I(t)$ and $P_{II}(r,t)$ are derived. In section 2.5.1 we give a heuristic derivation of the equations which provides an intuitive understanding of their relation to the underlying processes. In section 2.5.2 some results regarding the model behaviour are obtained through the analysis of the equations. Finally, a more rigorous derivation of the equations is given in section 2.5.3.

2.5.1 Dynamic equations

We define the following notation. Let $\mathbf{x}, \mathbf{y}$, etc. denote the positions of arbitrary sites on the lattice, and $f_{\mathbf{xy}} = f(r_{xy})$ denote the interaction between $\mathbf{x}$ and $\mathbf{y}$ at distance $r_{xy} = |\mathbf{x} - \mathbf{y}|$.

We assume that the system is spatially–homogeneous and isotropic at all times and has no boundaries, so that all sites are statistically equivalent. This is a common simplifying assumption (e.g. Bolker & Pacala, 1997); a more general version of these equations, valid also for heterogeneous systems, is given by (2.24) and (2.26). Spatial homogeneity implies that the probability that $\mathbf{x}$ is infected at time t, $\langle S_{\mathbf{x}}(t) \rangle$, is the same for all sites in the lattice and coincides with the expected fraction of infecteds $P_I(t)$:

$$\langle S_{\mathbf{x}}(t) \rangle = \langle S_{\mathbf{y}}(t) \rangle = P_I(t), \quad \forall\, \mathbf{x}, \mathbf{y} \in \text{system}, \tag{2.20}$$

and that the density $\langle S_{\mathbf{x}}(t) S_{\mathbf{y}}(t) \rangle$, i.e. the probability that sites $\mathbf{x}$ and $\mathbf{y}$ are both infected at time t, only depends on the distance r_{xy} between the two sites.

We shall show that $P_I(t)$ and $P_{II}(r_{xy}, t)$ obey the following differential equations:

$$\frac{dP_I(t)}{dt} = J_1 - (J_1 - J_2 + R)\ P_I(t) - J_2 \sum_{\substack{\mathbf{y} \in \ lattice \\ (\mathbf{y} \neq \mathbf{x})}} P_{II}(r_{xy}, t)\ f(r_{xy}) \tag{2.21}$$

and

$$\begin{aligned} \frac{1}{2}\frac{dP_{II}(r_{xy}, t)}{dt} = & \quad [J_1 + J_2\ f(r_{xy})]\ P_I(t) \\ & - (J_1 + J_2\ f(r_{xy}) + R)\ P_{II}(r_{xy}, t) \\ & + J_2 \sum_{\substack{\mathbf{z} \in lattice \\ (\mathbf{z} \neq \mathbf{x}, \mathbf{y})}} [P_{II}(r_{yz}, t) - P_{III}(r_{xy}, r_{xz}, r_{yz}, t)]\ f(r_{xz}), \end{aligned} \tag{2.22}$$

where $\mathbf{x}$ and $\mathbf{y}$ are arbitrary sites at distance r_{xy}, and $P_{III}(r_{xy}, r_{xz}, r_{yz}, t)$ is the probability that sites $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are all infected at time t. Since the system is spatially stationary, this probability only depends on the distances between

the sites r_{xy}, r_{xz} and r_{yz}. Equation (2.22) actually represents a whole system of equations, one for each distance r_{xy} between sites on the lattice.

Equations (2.21) and (2.22) show that P_I depends on P_{II}, and P_{II} depends on P_{III}. They are the first of an infinite hierarchy of equations in which the nth equation depends on the probability that $n+1$ sites are infected. Any sub–system of equations is not closed, as a direct consequence of the interactions in the system. In order to solve the system (2.21)–(2.22) we need to close it. This can be done assuming approximate relations between higher and lower–order probabilities. Approaches based on cluster approximations (Matsuda *et al.*, 1992; Levin & Durrett, 1997; Filipe & Gibson, 1998, 2000) can be applied to models with NN interactions (see chapter 4). Closure approximations suitable for general interactions will be the subject of chapter 6 (Filipe & Maule, 2000 b).

Following similar arguments to Filipe & Gibson (1998), we now show how these equations can be derived in a heuristic way.

We consider a generic site $\mathbf{x}$ on the lattice and the density $\langle S_{\mathbf{x}}(t)\rangle$, which represents the probability that $\mathbf{x}$ is infected at time t. The change in $\langle S_{\mathbf{x}}(t)\rangle$ during a small time interval $(t, t+dt)$ is given by

$$\begin{aligned} d\langle S_{\mathbf{x}}(t)\rangle = \; & \langle [1 - S_{\mathbf{x}}(t)] \text{ Prob}[0 \to 1; \mathbf{x}; t, dt]\rangle \\ - \; & \langle S_{\mathbf{x}}(t) \text{ Prob}[1 \to 0; \mathbf{x}; t, dt]\rangle, \end{aligned} \tag{2.23}$$

since a susceptible individual can acquire the disease with probability given by (2.1) and an infected can recover with probability given by (2.2). Substituting the conditional probabilities (2.1) and (2.2) in (2.23), we obtain

$$\begin{aligned} & \frac{d\langle S_{\mathbf{x}}(t)\rangle}{dt} = J_1 - (J_1 + R)\ \langle S_{\mathbf{x}}(t)\rangle + \\ & J_2 \sum_{\substack{\mathbf{y} \in \text{ lattice} \\ (\mathbf{y} \neq \mathbf{x})}} [\langle S_{\mathbf{y}}(t)\rangle - \langle S_{\mathbf{x}}(t) S_{\mathbf{y}}(t)\rangle]\ f_{\mathbf{xy}}, \end{aligned} \tag{2.24}$$

which reduces to equation (2.21) for a homogeneous unbounded system, for which $\langle S_{\mathbf{x}}(t)\rangle = P_I(t),\ \forall\ \mathbf{x} \in$ system.

We now consider the density $\langle S_{\mathbf{x}}(t) S_{\mathbf{y}}(t)\rangle$. Taking into account the possible changes during $(t, t+dt)$, and assuming that dt is sufficiently small to contain at most one event, we have

$$\begin{aligned} d\langle S_{\mathbf{x}}(t) S_{\mathbf{y}}(t)\rangle = \; & \langle S_{\mathbf{x}}(t)[1 - S_{\mathbf{y}}(t)] \text{ Prob}[1\ 0 \to 1\ 1; \mathbf{x}, \mathbf{y}; t, dt]\rangle \\ + \; & \langle [1 - S_{\mathbf{x}}(t)] S_{\mathbf{y}}(t) \text{ Prob}[0\ 1 \to 1\ 1; \mathbf{x}, \mathbf{y}; t, dt]\rangle \\ - \; & \langle S_{\mathbf{x}}(t) S_{\mathbf{y}}(t) \text{ Prob}[1\ 1 \to 1\ 0; \mathbf{x}, \mathbf{y}; t, dt]\rangle \\ - \; & \langle S_{\mathbf{x}}(t) S_{\mathbf{y}}(t) \text{ Prob}[1\ 1 \to 0\ 1; \mathbf{x}, \mathbf{y}; t, dt]\rangle, \end{aligned} \tag{2.25}$$

i.e. given a susceptible and an infected, the susceptible acquires the disease either from the infected in the pair or from any other infected in the lattice; in a pair of infecteds one is recovered. Again we use the transition probabilities (2.1) and (2.2) to write

$$\begin{aligned}\frac{d\langle S_{\mathbf{x}}(t)S_{\mathbf{y}}(t)\rangle}{dt} &= J_1[\langle S_{\mathbf{x}}(t)\rangle + \langle S_{\mathbf{y}}(t)\rangle] - 2\ (J_1+R)\ \langle S_{\mathbf{x}}(t)S_{\mathbf{y}}(t)\rangle \\ &+ J_2 \sum_{\mathbf{z}(\neq\mathbf{x})} [\langle S_{\mathbf{y}}(t)S_{\mathbf{v}}(t)\rangle - \langle S_{\mathbf{x}}(t)S_{\mathbf{y}}(t)S_{\mathbf{v}}(t)\rangle]\ f_{\mathbf{xz}} \\ &+ J_2 \sum_{\mathbf{z}(\neq\mathbf{y})} [\langle S_{\mathbf{x}}(t)S_{\mathbf{z}}(t)\rangle - \langle S_{\mathbf{x}}(t)S_{\mathbf{y}}(t)S_{\mathbf{z}}(t)\rangle]\ f_{\mathbf{yz}},\end{aligned} \tag{2.26}$$

which reduces to equation (2.22) for a homogeneous unbounded system.

2.5.2 Analytical results

Equations (2.21) and (2.22) can be used to study the behaviour of the model. We should recall that these equations describe the dynamics of a system which is spatially–homogeneous at all times.

Let us consider equation (2.21) and assume that there is no background source of infection ($J_1 = 0$):

$$\frac{dP_I(t)}{dt} = (J_2 - R)\ P_I(t) - J_2 \sum_{\substack{\mathbf{y}\in\ lattice \\ (\mathbf{y}\neq\mathbf{x})}} P_{II}(r_{xy},t)\ f_{\mathbf{xy}}. \tag{2.27}$$

From (2.17), we can write

$$P_{II}(r_{xy},t) = \sigma[S_{\mathbf{x}}(t)]\ \sigma[S_{\mathbf{y}}(t)]C_{\mathbf{xy}}(t) + \langle S_{\mathbf{x}}(t)\rangle\langle S_{\mathbf{y}}(t)\rangle, \tag{2.28}$$

which, in a spatially–homogeneous system reads

$$P_{II}(r_{xy},t) = (P_I - P_I^2)\ C_{\mathbf{xy}}(t) + P_I^2. \tag{2.29}$$

Equation (2.27) can then be rewritten as (Filipe *et al.*, 2000):

$$\frac{dP_I(t)}{dt} = -R\ P_I(t) + J_2 P_I(t)[1 - P_I(t)]\ [1 - \Gamma(t)], \tag{2.30}$$

where we have defined

$$\Gamma(t) \equiv \sum_{\mathbf{y}(\neq\mathbf{x})} C_{\mathbf{xy}}(t)\ f_{\mathbf{xy}} \approx \sum_{r=1}^{L} G(r)C(r,t)f(r), \tag{2.31}$$

and L is the linear size of the lattice.

This formulation clearly shows how spatial correlations produce a deviation of behaviour from the standard Mean Field (MF) approximation (e.g. Filipe &

Gibson, 1998). MF assumes that individuals are independent of each other and hence that there are no correlations ($\Gamma = 0$). The MF approximation exactly describes the model in the limit when $a = 0$, in which interactions are independent of the spatial location of individuals and no correlations develop. Within the MF approximation, equation (2.30) becomes

$$\frac{dP_I(t)}{dt} = P_I(t)\,[-R + J_2\,(1 - P_I(t))]. \tag{2.32}$$

Equation (2.32) can be solved explicitly (see, for example, Filipe & Gibson (1998) and chapter 4).

From (2.30) and (2.31) it is easy to see that if correlations are large or long–ranged the deviation from MF behaviour can be significant. A more detailed explanation of the MF approximation can be found in chapter 4.

In the stationary regime, the expected fraction of infecteds in an infinite system is constant, which means that its derivative with respect to time is zero. If we set $dP_I/dt = 0$, we can formally solve (2.30) and obtain an expression for the size of the epidemic at equilibrium:

$$P_I = 1 - \frac{R/J_2}{1 - \Gamma} \tag{2.33}$$

for $J_2/R > 1/(1 - \Gamma)$, and $P_I = 0$ otherwise. Hence the critical threshold value of J_2/R is $1/(1 - \Gamma)$, whereas the MF prediction for the threshold is 1 (Filipe *et al.*, 2000).

We now present the proof, given in Filipe *et al.* (2000), which demonstrates that the behaviour of P_I converges to the MF prediction and that the transition threshold is equal to 1 in the limit of an infinite system ($L \to \infty$) for all values of a between 0 and 2, where a is the power exponent of the contact distribution. In other words, we wish to show that

$$\lim_{L\to\infty} \Gamma = 0, \quad \forall\, a \;:\; 0 \le a \le 2, \tag{2.34}$$

whilst for $a > 2$ the limit is positive and constant. The only assumption we need to make is that spatial correlations do not decay more slowly than a power–law, i.e.

$$C(r) \le \frac{A}{r^b} \quad \text{for} \quad r \to \infty, \tag{2.35}$$

where $A,\ b > 0$.

Using (2.31), we write

$$\lim_{L\to\infty} \Gamma = \lim_{L\to\infty} \sum_{r=1}^{L} G(r)C(r,t)f(r). \tag{2.36}$$

The contact distribution (2.3) can be rewritten as

$$f(r) = \frac{r^{-a}}{\sum_{s=1}^{L} G(s)\ s^{-a}}, \tag{2.37}$$

where r and s are lattice distances, i.e. $r, s \in \{1, \sqrt{2}, 2, \sqrt{5}, \dots\}$. We now replace sums by integrals, which should be asymptotically accurate for the dominant (large r) part of the sums when $L \to \infty$, use $G(r) = 2\pi r$, and substitute (2.37) in to (2.36), to give

$$\lim_{L\to\infty} \Gamma = \lim_{L\to\infty} \frac{\int_1^L C(r)\ r^{(1-a)} dr}{\int_1^L r^{(1-a)} dr}. \tag{2.38}$$

Using assumption (2.35), we can then write

$$\lim_{L\to\infty} \Gamma \leq \lim_{L\to\infty} \frac{\int_1^L r^{(1-a-b)} dr}{\int_1^L r^{(1-a)} dr}. \tag{2.39}$$

To calculate this integral we distinguish three cases.

Case 1: $a \neq 2$, $a + b \neq 2$

Equation (2.39) reads

$$\lim_{L\to\infty} \Gamma \leq \lim_{L\to\infty} \frac{(2-a)(L^{(2-a-b)} - k_1)}{(2-a-b)(L^{(2-a)} - k_2)}, \tag{2.40}$$

where k_1, k_2 are positive constants. If $a > 2$ the limit is just a positive constant; if $a < 2$, $L^{(2-a)}$ tends to infinity, hence we have two cases:

(1) $a + b > 2$:

$$\Gamma \sim \frac{k_1'}{L^{(2-a)}} \to 0; \tag{2.41}$$

(2) $a + b < 2$:

$$\Gamma \sim L^{-b} \to 0; \tag{2.42}$$

Case 2: $a = 2$, $(a + b > 2)$

Equation (2.39) reads

$$\lim_{L\to\infty} \Gamma \leq \lim_{L\to\infty} \frac{-(1/b)(L^{-b} - c_1)}{\ln L} = 0; \tag{2.43}$$

Case 3: $a < 2$, $a + b = 2$
Equation (2.39) reads

$$\lim_{L\to\infty} \Gamma \leq \lim_{L\to\infty} \frac{(2-a)\ln L}{L^{(2-a)} - c_2} = 0, \tag{2.44}$$

since the denominators diverge more quickly than the numerators (c_1, c_2 are constants).

We have proved that, for an infinite system, Γ vanishes for all values of a smaller than 2, hence MF predictions for the behaviour of P_I and the transition threshold become exact (Filipe *et al.*, 2000). This proof easily extends to any spatial dimensions d, so that the result can be generalised to $\Gamma \underset{L\to\infty}{\longrightarrow} 0$ for $a \leq d$. A similar statement for physical systems has also been conjectured (Cannas & Tamarit, 1996; Cannas, 1998). Note that this result does not imply that correlations are identically zero for $a \leq 2$ (a point supported by the simulation results in the next chapter), which means that only some aspects of epidemic behaviour are described by the MF limit.

2.5.3 Formal derivation of dynamic equations

This section presents a mathematically rigorous derivation of equations (2.21) and (2.22). We now consider the more general case in which the system is not assumed to be spatially–homogeneous. We shall derive equations for the evolution of the expectation of binary variables corresponding to specific sites, i.e. the equation for $\langle S_{\mathbf{x}}(t) S_{\mathbf{y}}(t)\rangle$ only refers to the specific individuals located at $\mathbf{x}$ and $\mathbf{y}$ and not to any pair of individuals at distance $|\mathbf{x} - \mathbf{y}|$ apart. Equations (2.21) and (2.22) represent a special case which can be obtained by assuming spatial invariance.

In this section we shall adopt $\mathbf{E}[\cdots]$ to denote expectation over stochastic realisations. This is the standard notation used in probability theory (see, for example, Grimmet & Stirzaker, 1992) and is equivalent to the one we have adopted so far: $\langle\cdots\rangle$.

First we derive a differential equation for the evolution in time of the expectation of the binary variable $S_{\mathbf{x}}(t)$, which describes the state of the individual located at site $\mathbf{x}$ on the lattice:

$$\mathbf{E}[S_{\mathbf{x}}(t)] = \sum_{j\in\{0,1\}} j\ \mathbf{P}[S_{\mathbf{x}}(t) = j] \ = \ \mathbf{P}[S_{\mathbf{x}}(t) = 1], \tag{2.45}$$

where $\mathbf{P}[S_{\mathbf{x}}(t) = j]$ is the probability that $S_{\mathbf{x}}$ has value j at time t. We consider

$$\mathbf{E}[S_{\mathbf{x}}(t + dt)] = \mathbf{E}[S_{\mathbf{x}}(t)] + \mathcal{A}\ dt. \tag{2.46}$$

From (2.45) and (2.46), we can write $\mathcal{A}\ dt$ as

$$\mathcal{A}\ dt = \mathbf{P}[S_{\mathbf{x}}(t+dt) = 1] - \mathbf{P}[S_{\mathbf{x}}(t) = 1]. \tag{2.47}$$

Using the property

$$\mathbf{P}[A] = \sum_{i=1}^{n} \mathbf{P}[A|B_i]\mathbf{P}[B_i], \tag{2.48}$$

where $B_1, B_2, \ldots, B_n$ is a partition of the sample space, we can rewrite (2.47) as

$$\begin{aligned}
&\sum_{\{S_{\mathbf{y}}(t)=i_{\mathbf{y}}:\mathbf{y}\neq\mathbf{x}\}} \mathbf{P}[S_{\mathbf{x}}(t+dt) = 1|S_{\mathbf{x}}(t) = 1, \{S_{\mathbf{y}}(t) = i_{\mathbf{y}} : \mathbf{y} \neq \mathbf{x}\}] \\
\times\ &\mathbf{P}[S_{\mathbf{x}}(t) = 1, \{S_{\mathbf{y}}(t) = i_{\mathbf{y}} : \mathbf{y} \neq \mathbf{x}\}] + \\
&\sum_{\{S_{\mathbf{y}}(t)=i_{\mathbf{y}}:\mathbf{y}\neq\mathbf{x}\}} \mathbf{P}[S_{\mathbf{x}}(t+dt) = 1|S_{\mathbf{x}}(t) = 0, \{S_{\mathbf{y}}(t) = i_{\mathbf{y}} : \mathbf{y} \neq \mathbf{x}\}] \\
\times\ &\mathbf{P}[S_{\mathbf{x}}(t) = 0, \{S_{\mathbf{y}}(t) = i_{\mathbf{y}} : \mathbf{y} \neq \mathbf{x}\}] - \\
&\mathbf{P}[S_{\mathbf{x}}(t) = 1],
\end{aligned} \tag{2.49}$$

where $i_{\mathbf{y}} \in \{0, 1\}$. In (2.49) we have rewritten $\mathbf{P}[S_{\mathbf{x}}(t+dt) = 1]$ as the probability that the individual $\mathbf{x}$ is infected at time $(t+dt)$, conditional on the possible states of all the other individuals at time t, times the corresponding probabilities.

We now consider the random variable

$$\mathbf{E}[(1 - S_{\mathbf{x}}(t))\ \mathbf{P}[S_{\mathbf{x}}(t+dt) = 1|S_{\mathbf{x}}(t), \{S_{\mathbf{y}}(t) = i_{\mathbf{y}} : \mathbf{y} \neq \mathbf{x}\}]]. \tag{2.50}$$

Note that (2.50) is a random variable because the probability is conditional on $S_{\mathbf{x}}(t)$, which is a random variable. For simplicity, in the following we let $\{\xi\}_{\mathbf{y}}$ denote $\{S_{\mathbf{y}}(t) = i_{\mathbf{y}} : \mathbf{y} \neq \mathbf{x}\}$. Then (2.50) is equal to

$$\begin{aligned}
&\sum_{\substack{\{\xi\}_{\mathbf{y}} \\ (S_{\mathbf{x}}(t)=0)}} 1 \times\ \mathbf{P}[S_{\mathbf{x}}(t+dt) = 1|S_{\mathbf{x}}(t) = 0, \{\xi\}_{\mathbf{y}}]\ \mathbf{P}[S_{\mathbf{x}}(t) = 0, \{\xi\}_{\mathbf{y}}] + \\
&\sum_{\substack{\{\xi\}_{\mathbf{y}} \\ (S_{\mathbf{x}}(t)=1)}} 0 \times\ \mathbf{P}[S_{\mathbf{x}}(t+dt) = 1|S_{\mathbf{x}}(t) = 1, \{\xi\}_{\mathbf{y}}]\ \mathbf{P}[S_{\mathbf{x}}(t) = 1, \{\xi\}_{\mathbf{y}}].
\end{aligned} \tag{2.51}$$

Analogously, we consider the random variable

$$\mathbf{E}[S_{\mathbf{x}}(t)\ \mathbf{P}[S_{\mathbf{x}}(t+dt) = 0|S_{\mathbf{x}}(t), \{\xi\}_{\mathbf{y}}]], \tag{2.52}$$

which is equal to

$$\sum_{\substack{\{\xi\}_{\mathbf{y}} \\ (S_{\mathbf{x}}(t)=1)}} \mathbf{P}[S_{\mathbf{x}}(t+dt) = 0|S_{\mathbf{x}}(t) = 1, \{\xi\}_{\mathbf{y}}]\ \mathbf{P}[S_{\mathbf{x}}(t) = 1, \{\xi\}_{\mathbf{y}}]. \tag{2.53}$$

Using the property of conditional probability

$$\mathbf{P}[A|B] + \mathbf{P}[A^c|B] = 1, \tag{2.54}$$

we can rewrite (2.53) as

$$\sum_{\substack{\{\xi\}_\mathbf{y} \\ (S_\mathbf{x}(t)=1)}} (1 - \mathbf{P}[S_\mathbf{x}(t+dt) = 1|S_\mathbf{x}(t) = 1, \{\xi\}_\mathbf{y}])\ \mathbf{P}[S_\mathbf{x}(t) = 1, \{\xi\}_\mathbf{y}]. \tag{2.55}$$

Using (2.50), (2.52) and (2.55), we can write (2.47) as

$$\begin{aligned} \mathcal{A}\ dt = \ & \mathbf{E}[(1 - S_\mathbf{x}(t))\ \mathbf{P}[S_\mathbf{x}(t+dt) = 1|S_\mathbf{x}(t), \{\xi\}_\mathbf{y}]] \\ - \ & \mathbf{E}[S_\mathbf{x}(t)\ \mathbf{P}[S_\mathbf{x}(t+dt) = 0|S_\mathbf{x}(t), \{\xi\}_\mathbf{y}]]. \end{aligned} \tag{2.56}$$

From the stochastic transition rules (2.1) and (2.2), we have

$$\mathbf{P}\left[S_\mathbf{x}(t+dt) = 1|S_\mathbf{x}(t) = 0, \{\xi\}_\mathbf{y}\right] = [J_1 + J_2 \sum_{\mathbf{y}(\neq \mathbf{x})} S_\mathbf{y}(t) f_{\mathbf{xy}}]\ dt \equiv z_1, \tag{2.57}$$

and

$$\mathbf{P}\left[S_\mathbf{x}(t+dt) = 0|S_\mathbf{x}(t) = 1, \{\xi\}_\mathbf{y}\right] = R\ dt \equiv z_2. \tag{2.58}$$

Hence:

$$\mathbf{P}\left[S_\mathbf{x}(t+dt) = 1|S_\mathbf{x}(t) = 1, \{\xi\}_\mathbf{y}\right] = 1 - z_1 \tag{2.59}$$

and

$$\mathbf{P}\left[S_\mathbf{x}(t+dt) = 0|S_\mathbf{x}(t) = 0, \{\xi\}_\mathbf{y}\right] = 1 - z_2. \tag{2.60}$$

We now consider the stochastic variable $Z = \mathbf{P}\left[S_\mathbf{x}(t+dt) = 1|S_\mathbf{x}(t), \{\xi\}_\mathbf{y}\right]$. This can be written as a linear combination of indicator variables (Grimmet & Stirzaker, 1992):

$$Z = \chi_{\{Z=z_1\}}\ z_1\ + \chi_{\{Z=(1-z_1)\}}\ (1 - z_1), \tag{2.61}$$

where

$$\chi_{\{Z=z_1\}} \begin{cases} 1 & \text{if } Z = z_1 \\ 0 & \text{if } Z \neq z_1. \end{cases} \tag{2.62}$$

Since Z can take value z_1 if, and only if, the variable $S_\mathbf{x}$ is 0 at time t, we have

$$\begin{aligned} \chi_{\{Z=z_1\}} = \ & \chi_{\{S_\mathbf{x}(t)=0\}} = 1 - S_\mathbf{x}(t) \\ \chi_{\{Z=(1-z_1)\}} = \ & \chi_{\{S_\mathbf{x}(t)=1\}} = S_\mathbf{x}(t) \end{aligned} \tag{2.63}$$

Thus finally we obtain

$$Z = [1 - S_{\mathbf{x}}(t)]\ z_1\ + S_{\mathbf{x}}(t)\ (1 - z_1). \tag{2.64}$$

Analogously,

$$\mathbf{P}\left[S_{\mathbf{x}}(t+dt) = 0 | S_{\mathbf{x}}(t), \{\xi\}_{\mathbf{y}}\right] = [1 - S_{\mathbf{x}}(t)]\ (1 - z_2)\ + S_{\mathbf{x}}(t)\ z_2. \tag{2.65}$$

Substituting (2.64) and (2.65) in (2.56), and recalling that $S_{\mathbf{x}}$ is a Bernoulli variable ($S_{\mathbf{x}}^2 = S_{\mathbf{x}}$), we obtain

$$\begin{aligned} \mathcal{A} = & \quad J_1 - (J_1 + R)\ \mathbf{E}[S_{\mathbf{x}}(t)] \\ & + \ J_2 \sum_{\mathbf{y}(\neq \mathbf{x})} (\mathbf{E}[S_{\mathbf{y}}(t)] - \mathbf{E}[S_{\mathbf{x}}(t) S_{\mathbf{y}}(t)])\ f_{\mathbf{xy}}. \end{aligned} \tag{2.66}$$

From (2.46) we have

$$\mathcal{A} = \frac{\mathbf{E}[S_{\mathbf{x}}(t+dt)] - \mathbf{E}[S_{\mathbf{x}}(t)]}{dt} \tag{2.67}$$

and in the limit $dt \to 0$

$$\mathcal{A} = \frac{d\mathbf{E}[S_{\mathbf{x}}(t)]}{dt}, \tag{2.68}$$

hence

$$\begin{aligned} \frac{d\mathbf{E}[S_{\mathbf{x}}(t)]}{dt} = & \quad J_1 - (J_1 + R)\ \mathbf{E}[S_{\mathbf{x}}(t)] \\ & + \ J_2 \sum_{\mathbf{y}(\neq \mathbf{x})} (\mathbf{E}[S_{\mathbf{y}}(t)] - \mathbf{E}[S_{\mathbf{x}}(t) S_{\mathbf{y}}(t)])\ f_{\mathbf{xy}}. \end{aligned} \tag{2.69}$$

Equation (2.21) is a particular case of (2.69) for spatially–homogeneous systems.

The dynamic equation for the density $\mathbf{E}[S_{\mathbf{x}}(t) S_{\mathbf{y}}(t)]$ can be derived in an analogous way.

$$\mathbf{E}[S_{\mathbf{x}}(t) S_{\mathbf{y}}(t)] = \sum_{j \in \{0,1\}} j\ \mathbf{P}[S_{\mathbf{x}}(t) S_{\mathbf{y}}(t) = j] \ = \ \mathbf{P}[S_{\mathbf{x}}(t) = 1, S_{\mathbf{y}}(t) = 1]. \tag{2.70}$$

We consider

$$\mathcal{B} = \mathbf{P}[S_{\mathbf{x}}(t+dt) = 1, S_{\mathbf{y}}(t+dt) = 1] - \mathbf{P}[S_{\mathbf{x}}(t) = 1, S_{\mathbf{y}}(t) = 1]. \tag{2.71}$$

Now we let $\{\xi\}_{\mathbf{z}}$ denote $\{S_{\mathbf{z}}(t) = i_{\mathbf{z}} : \mathbf{z} \neq \mathbf{x}, \mathbf{y}\}$.

Using the above argument, (2.71) can be written as

$$
\begin{aligned}
&\sum_{\{\xi\}_{\mathbf{z}}} \mathbf{P}[S_{\mathbf{x}}(t+dt)=1, S_{\mathbf{y}}(t+dt)=1 | S_{\mathbf{x}}(t)=0, S_{\mathbf{y}}(t)=1, \{\xi\}_{\mathbf{z}}] \\
\times\ &\mathbf{P}[S_{\mathbf{x}}(t)=0, S_{\mathbf{y}}(t)=1, \{\xi\}_{\mathbf{z}}] + \\
&\sum_{\{\xi\}_{\mathbf{z}}} \mathbf{P}[S_{\mathbf{x}}(t+dt)=1, S_{\mathbf{y}}(t+dt)=1 | S_{\mathbf{x}}(t)=1, S_{\mathbf{y}}(t)=0, \{\xi\}_{\mathbf{z}}] \\
\times\ &\mathbf{P}[S_{\mathbf{x}}(t)=1, S_{\mathbf{y}}(t)=0, \{\xi\}_{\mathbf{z}}] + \\
&\sum_{\{\xi\}_{\mathbf{z}}} \mathbf{P}[S_{\mathbf{x}}(t+dt)=1, S_{\mathbf{y}}(t+dt)=1 | S_{\mathbf{x}}(t)=1, S_{\mathbf{y}}(t)=1, \{\xi\}_{\mathbf{z}}] \\
\times\ &\mathbf{P}[S_{\mathbf{x}}(t)=1, S_{\mathbf{y}}(t)=1, \{\xi\}_{\mathbf{z}}] + \\
&\sum_{\{\xi\}_{\mathbf{z}}} \mathbf{P}[S_{\mathbf{x}}(t+dt)=1, S_{\mathbf{y}}(t+dt)=1 | S_{\mathbf{x}}(t)=0, S_{\mathbf{y}}(t)=0, \{\xi\}_{\mathbf{z}}] \\
\times\ &\mathbf{P}[S_{\mathbf{x}}(t)=0, S_{\mathbf{y}}(t)=0, \{\xi\}_{\mathbf{z}}] \\
&-\mathbf{P}[S_{\mathbf{x}}(t)=1, S_{\mathbf{y}}(t)=1].
\end{aligned}
\tag{2.72}
$$

The probability of two events occurring in the time interval $(t, t+dt)$ is of order $O(dt^2)$; hence the fourth summation in (2.72) is negligible as it contains the probability of both individuals at $\mathbf{x}$ and $\mathbf{y}$ being infected at time $(t+dt)$, given that they were susceptible at time t. Equation (2.72) can be rewritten in terms of conditional expectations:

$$
\begin{aligned}
&\mathbf{E}[(1-S_{\mathbf{x}}(t))S_{\mathbf{y}}(t)\ \mathbf{P}[S_{\mathbf{x}}(t+dt)=1, S_{\mathbf{y}}(t+dt)=1 | S_{\mathbf{x}}(t), S_{\mathbf{y}}(t), \{\xi\}_{\mathbf{z}}]] \\
+\ &\mathbf{E}[S_{\mathbf{x}}(t)(1-S_{\mathbf{y}}(t))\ \mathbf{P}[S_{\mathbf{x}}(t+dt)=1, S_{\mathbf{y}}(t+dt)=1 | S_{\mathbf{x}}(t), S_{\mathbf{y}}(t), \{\xi\}_{\mathbf{z}}]] \\
-\ &\mathbf{E}[S_{\mathbf{x}}(t)S_{\mathbf{y}}(t)\ \mathbf{P}[S_{\mathbf{x}}(t+dt)=1, S_{\mathbf{y}}(t+dt)=0 | S_{\mathbf{x}}(t), S_{\mathbf{y}}(t), \{\xi\}_{\mathbf{z}}]] \\
-\ &\mathbf{E}[S_{\mathbf{x}}(t)S_{\mathbf{y}}(t)\ \mathbf{P}[S_{\mathbf{x}}(t+dt)=0, S_{\mathbf{y}}(t+dt)=1 | S_{\mathbf{x}}(t), S_{\mathbf{y}}(t), \{\xi\}_{\mathbf{z}}]].
\end{aligned}
\tag{2.73}
$$

We cannot substitute the stochastic transition rules (2.1) and (2.2) in (2.73) as we did before because we are now dealing with joint probabilities for both individuals at $\mathbf{x}$ and $\mathbf{y}$ at time $(t+dt)$. However, using the property

$$
\mathbf{P}[A|C] = \mathbf{P}[(A \cap B) \cup (A \cap B^c)|C] = \mathbf{P}[A \cap B|C] + \mathbf{P}[A \cap B^c|C], \tag{2.74}
$$

we can write, for example,

$$
\begin{aligned}
&\mathbf{P}[S_{\mathbf{x}}(t+dt)=1, S_{\mathbf{y}}(t+dt)=1 | S_{\mathbf{x}}(t)=0, S_{\mathbf{y}}(t)=1, \{\xi\}_{\mathbf{z}}] = \\
&\mathbf{P}[S_{\mathbf{x}}(t+dt)=1 | S_{\mathbf{x}}(t)=0, S_{\mathbf{y}}(t)=1, \{\xi\}_{\mathbf{z}}] - \\
&\mathbf{P}[S_{\mathbf{x}}(t+dt)=1, S_{\mathbf{y}}(t+dt)=0 | S_{\mathbf{x}}(t)=0, S_{\mathbf{y}}(t)=1, \{\xi\}_{\mathbf{z}}],
\end{aligned}
\tag{2.75}
$$

in which the last term is of order $O(dt^2)$ and hence negligible. This argument applies to all the conditional probabilities in (2.73). Note that the transition probability from susceptible to infected for individual $\mathbf{x}$, (2.1), given that individual

$\mathbf{y}$ is infected ($S_{\mathbf{y}}(t) = 1$), can be rewritten as

$$\mathbf{P}[S_{\mathbf{x}}(t+dt) = 1 | S_{\mathbf{x}}(t) = 0, S_{\mathbf{y}}(t) = 1, \{\xi\}_{\mathbf{z}}] = [J_1 + J_2\, f_{\mathbf{xy}} + J_2 \sum_{\mathbf{z}(\neq \mathbf{x},\mathbf{y})} f_{\mathbf{xy}} S_{\mathbf{z}}(t)]\, dt. \tag{2.76}$$

Taking the limit $dt \to 0$, and using (2.73), (2.76) and (2.2), we finally obtain

$$\frac{d\mathbf{E}[S_{\mathbf{x}}(t) S_{\mathbf{y}}(t)]}{dt} = J_1 (\mathbf{E}[S_{\mathbf{x}}(t)] + \mathbf{E}[S_{\mathbf{y}}(t)]) \tag{2.77}$$

$$- \; 2\,(J_1 + J_2\, f_{\mathbf{xy}} + R)\; \mathbf{E}[S_{\mathbf{x}}(t) S_{\mathbf{y}}(t)] + J_2\; (\mathbf{E}[S_{\mathbf{x}}(t)] + \mathbf{E}[S_{\mathbf{y}}(t)])\; f_{\mathbf{xy}}$$

$$+ \quad J_2 \sum_{\mathbf{z}(\neq \mathbf{x},\mathbf{y})} (\mathbf{E}[S_{\mathbf{y}}(t) S_{\mathbf{z}}(t)] - \mathbf{E}[S_{\mathbf{x}}(t) S_{\mathbf{y}}(t) S_{\mathbf{z}}(t)])\; f_{\mathbf{xz}}$$

$$+ \quad J_2 \sum_{\mathbf{z}(\neq \mathbf{x},\mathbf{y})} (\mathbf{E}[S_{\mathbf{x}}(t) S_{\mathbf{z}}(t)] - \mathbf{E}[S_{\mathbf{x}}(t) S_{\mathbf{y}}(t) S_{\mathbf{z}}(t)])\; f_{\mathbf{yz}}.$$

Equation (2.77) represents a system of equations for each and every pair of lattice sites $\mathbf{x}$ and $\mathbf{y}$. Equation (2.22) is a particular case of (2.77) for spatially–homogeneous systems.

The same arguments used to derive the second set of equations in the infinite hierarchy can be used to derive the nth set of equations of the hierarchy:

$$d\mathbf{E}[S_{\mathbf{x}_1}(t) \cdots S_{\mathbf{x}_n}] = (\sum_{i=1}^{n} \mathbf{E}[\; (1 - S_{\mathbf{x}_i}(t)) \prod_{\substack{k=1 \\ (k \neq i)}}^{n} S_{\mathbf{x}_k}(t)$$

$$\times \quad [J_1 + J_2 \sum_{\substack{j=1 \\ (j \neq i)}}^{n} f_{\mathbf{x}_i \mathbf{x}_j} + J_2 \sum_{\mathbf{v}(\neq \mathbf{x}_1, \ldots, \mathbf{x}_n)} f_{\mathbf{x}_i \mathbf{v}} S_{\mathbf{v}}(t)]\;]) \tag{2.78}$$

$$- \qquad \mathbf{E}[S_{\mathbf{x}_1} \ldots S_{\mathbf{x}_n}]\; n\; R.$$

The system of equations (2.69)–(2.78) is not closed for any finite n and therefore cannot be solved. The special case represented by spatially–homogeneous systems can be recovered by imposing $\mathbf{E}[S_{\mathbf{x}}] = \mathbf{E}[S_{\mathbf{y}}], \forall\; \mathbf{x}, \mathbf{y} \in$ system.

2.6 Summary

A lattice model for the development of plant epidemics has been presented. The dynamics of the infection process are determined by three parameters: the primary and secondary infection rates (J_1 and J_2) and the power exponent of the contact distribution, a, which specifies the range of disease dispersal. The contact distribution decays as a power–law of distance and can describe interactions ranging from local contacts between neighbours to uniform interactions, independent of the relative position of individuals. A recovery process, whereby infecteds are replaced by susceptibles, can be included, which results in two possible versions

of the model, namely the SI model (without recovery) and the SIS model (with recovery). The model is simple but incorporates important realistic features, such as stochasticity and spatial extension. Moreover, its simplicity allows us to give a clear interpretation of the parameters which affect its behaviour.

Close attention was paid to the spatial characteristics of the model because, when studying the model behaviour, we wish to understand how different mechanisms of dispersal affect the development of epidemics, what kind of disease patterns they produce, and how the resulting spatial distribution of infecteds in turn influences the persistence or extinction of epidemics. For this purpose we defined the quantities which will be measured in the simulation of the model and described how they will be estimated using samples of many stochastic realisations. In particular, we are interested in the size of the epidemic and the relation between the number and the spatial distribution of infecteds.

We also derived a set of individual–based dynamic equations describing the expected dynamics of the spatial and stochastic model. These equations represent a valuable alternative to computationally–intensive stochastic simulation and can be used to obtain predictions for the expected behaviour. The problem with stochastic and spatial models is that they are described by an infinite hierarchy of equations which is mathematically intractable. Yet from a simple analysis of the first two equations in the hierarchy, we established some properties of the model and obtained theoretical predictions for its behaviour in particular situations. Closure approximations will be used to close the system and solve the equations, and these solutions will be tested against simulation (see chapter 6, Filipe & Maule, 2000 b). These equations represent the starting point for the development of a simple and robust analytical framework for studying the behaviour of a wide range of models characterised by general interactions.

In conclusion, we have two possible formulations for the study of our epidemic model, one based on a stochastic approach and the other on an analytical approach. The stochastic approach requires the use of simulation; it is realistic, since it accounts for the uncertainty affecting natural systems, but leads to findings which are difficult to generalise to other models. The analytical approach is less accurate since it relies on approximate methods, but on the other hand is very general and can be used for investigating a large variety of interacting particle systems.

Chapter 3

Simulation of the spatio–temporal stochastic model

3.1 Introduction

In recent years, the standard division of subjects such as physics, chemistry and biology into "experimental" and "theoretical" has been eclipsed by the advent of computer simulations, which represent a third and powerful approach. Computer simulations provide a method for investigating the behaviour of models which are not analytically solvable and for which experiments or collections of field data cannot provide conclusive information (Ripley, 1987; Binder & Heermann, 1992).

Computer simulations yield exact information (apart from statistical errors, which in principle can be made as small as desired) about model systems which are precisely characterised. In much the same way as the output from experimental trials may vary, so too can the output from stochastic realisations of a simulated process which is aimed at mimicking the variability observed in the former. However, computer simulations have the advantage of allowing us to control all the variables which may influence the behaviour of the system. Moreover, simulations can provide information that would be very difficult to obtain directly from experiments.

In order to study the average behaviour of a system, a very large number of realisations may be needed to obtain a meaningful estimation of quantities subject to large stochastic variation. This is easily done using computer simulations, whilst repeated realisations of the same experiment under the same conditions may prove to be a difficult task. On the other hand, however appealing, the study of model behaviour using analytical approaches can also be difficult and, in the majority of cases, requires the use of approximations with uncontrollable effects.

A stochastic spatio–temporal model for the spread of infectious diseases in a lattice–distributed population of plants is mathematically intractable. In the

previous chapter we saw that the differential equations describing the dynamics of an epidemic form an infinite system in which any sub–system is not close. Nevertheless, a model which incorporates the discreteness of individuals, spatial extension and stochasticity provides a realistic representation of the biological mechanisms occurring in real systems. Computer simulations are a valuable tool for studying the behaviour of such models. They provide insight into the underlying mechanisms, highlight unforeseen features of a process, and suggest further profitable lines of biological investigation.

In this chapter we present the results of a thorough simulation study of the stationary and transient behaviour of the model introduced in chapter 2. Some of these results are reported in Filipe *et al.*, 2000 and Filipe & Maule, 2000 a. In the stationary regime, the relation between epidemic size and disease distribution, characterised by spatial autocorrelations, and their dependence on dispersal and infectiousness parameters, is demonstrated. Special attention is given to boundary effects, which can significantly decrease disease levels relative to standard periodic geometries in the cases of long and medium–distance dispersal. In the transient regime we attempt to characterise transients in terms of the mean and variance of a distribution of disease progress curves, and to identify universal and specific features of behaviour. In particular, we propose and test a definition of average transient duration which captures well the dependence on model parameters. The time–evolution of spatial autocorrelations reveals distinctive features (such as power–law distance decay) between stationary and transient behaviour. The effects of having no recovery or having different initial conditions are also examined.

3.2 Stochastic algorithm

For a given set of parameters J_1, J_2 and a, we simulate a stochastic realisation of the model by specifying the time, nature (infection or recovery) and location of the events. Let us suppose we know the state of each site in the lattice at time t, and let $N_S(t)$ and $N_I(t)$ be the total number of susceptible and infected individuals ($N_S(t) + N_I(t) = N$). The time until next event, s, is simulated from an exponential distribution with mean $1/\Sigma(t)$ (Renshaw, 1991; Grimmett & Stirzaker, 1992), where

$$\Sigma(t) = RN_I(t) + J_1 N_S(t) + J_2 \sum_{\mathbf{x}} [1 - S_{\mathbf{x}}(t)] \sum_{\mathbf{y} \neq \mathbf{x}} f(|\mathbf{x} - \mathbf{y}|) S_{\mathbf{y}}(t). \qquad (3.1)$$

$\Sigma(t)\ dt$ is the probability that any event occurs in a small time interval $(t, t + dt)$. $\Sigma(t)$ depends only on the actual state of the system at time t and not on

the previous history of the system, i.e. this is a Markov process (Grimmett & Stirzaker, 1992). Thus the current time t is then incremented by s.

Next, the nature of the event occurring at $t+s$ is determined. With probability $R\ N_I(t)/\Sigma(t)$, this is selected to be the recovery of an infected site, and a particular site is chosen randomly from the set of infecteds and replaced by a susceptible. If the selected event is the infection of a susceptible, which has probability $[J_1 N_S(t) + J_2 \sum_{\mathbf{x}}[1-S_{\mathbf{x}}(t)] \sum_{\mathbf{y}\neq\mathbf{x}} f(|\mathbf{x}-\mathbf{y}|)S_{\mathbf{y}}(t)]/\Sigma(t)$, a site $\mathbf{x}$ is chosen from the set of all susceptibles with probability proportional to the infective challenge experienced by $\mathbf{x}$, that is

$$\Phi_{\mathbf{x}}(t) = J_1 + J_2 \sum_{\mathbf{y}\neq\mathbf{x}} f(|\mathbf{x}-\mathbf{y}|)S_{\mathbf{y}}(t), \tag{3.2}$$

and its state changed to infected. By repeating this procedure from a set of initial conditions, we generate a stochastic realisation of the model.

3.3 Numerical implementation

Computer simulations were carried out on SUN Ultra 2170 and Ultra 1140 Workstations. In situations where a very large number of stochastic realisations were necessary, simulations were run in parallel on a 512-alpha Cray Research T3D at the Edinburgh Parallel Computing Centre (EPCC), using the Message Passing Interface standard, version 1.1 (EPCC, 1995).

The numerical algorithm was implemented in Fortran 77 and Fortran 90. The random number generator used was the intrinsic Fortran routine *ran*, which generates random numbers with a uniform distribution using a multiplicative congruential algorithm. Tests were made on the performance of the random number generator used (Filipe, private communication), and the results were found to be satisfactory for our needs.

3.4 Parameter choice

Since each susceptible can interact with any infected in the population, stochastic realisations are computationally intensive, especially for large systems, and the model behaviour may depend significantly on the size of the system. Real populations exhibit system–size effects of two kinds: finite–size statistical variation, and edge effects that depend on specific boundaries. In order to make results general and informative we wish to minimise these effects. This can be done by simulating the model on a relatively large lattice ($L = 200$, corresponding to $N = 40,000$ individuals) without boundaries, whose average behaviour we expect

to be close to that of an infinitely–large system. To have some idea of the differences between this paradigm and a real system we compared average behaviour in lattices of size $L = 100$ and 200, with and without boundaries (FBC and PBC). Finite–size stochastic variation within and between realisations was substantially reduced from $L = 100$ to 200, which facilitated sampling.

For the stationary, ergodic regime, the size of simulated systems is chosen to be $L = 200$. Typical samples consist of ~ 10 realisations. In order to reduce computational cost, smaller systems, of size $L = 100$, are adopted in the transient regime, where samples consist of hundreds of realisations (128 to 256, performed on a parallel computer). Large samples are necessary in this case since it is not possible to average over time. The choice of a smaller system size is appropriate also because part of the transient variability exhibited by biological populations, for example in replicate experiments (Kleczkowski *et al.*, 1996), results from their relatively small size. $L = 100$ seems to be small enough to sustain a certain degree of variability yet large enough to allow qualitative conclusions to be independent of L.

We focussed on values $a = 2, 3.5, 20$ for the dispersal exponent, representing cases of long–range (LR), medium–range (MR) and short–range (SR) dispersal.

One effect of background or primary infection on disease progress is to randomise the disease pattern. A much less understood aspect of epidemic development is how the self–generated distribution of infecteds conditions disease transmission. In the absence of primary infection the system exhibits a phase transition from epidemic persistence to extinction. For clarity, we concentrate on this case and set $J_1 = 0$. Some examples of epidemics generated by both sources of infection are shown in later chapters, but even then background infection has low levels in order to avoid the dominance of the random process over the interaction process.

A wide range of values of J_2 are explored in each dispersal case: from the persistence threshold (the value below which the epidemic always becomes extinct) to values corresponding to large long–term epidemic size ($\geq 80\%$ of hosts infected).

Epidemics are initiated with a small fraction of randomly distributed infecteds. Stationary behaviour is independent of the initial condition, provided that a sufficiently long time has elapsed from the start of the observation. For the study of the transient regimes, the initial number of infecteds is small, typically 20 infecteds in a population of 10,000 individuals.

In this chapter, $\mu \equiv P_I$ is an abbreviation for the expected epidemic size, and $\beta \equiv J_2/R$ is the relative rate of secondary infection which we call the infectious-

ness.

3.5 Stationary behaviour

First we analyse the stationary behaviour of the model which includes recovery of infecteds. Lattices with and without boundaries (FBC and PBC) are considered.

3.5.1 Epidemic size

Fig. 3.1 displays the expected epidemic size μ against infectiousness β for dispersal

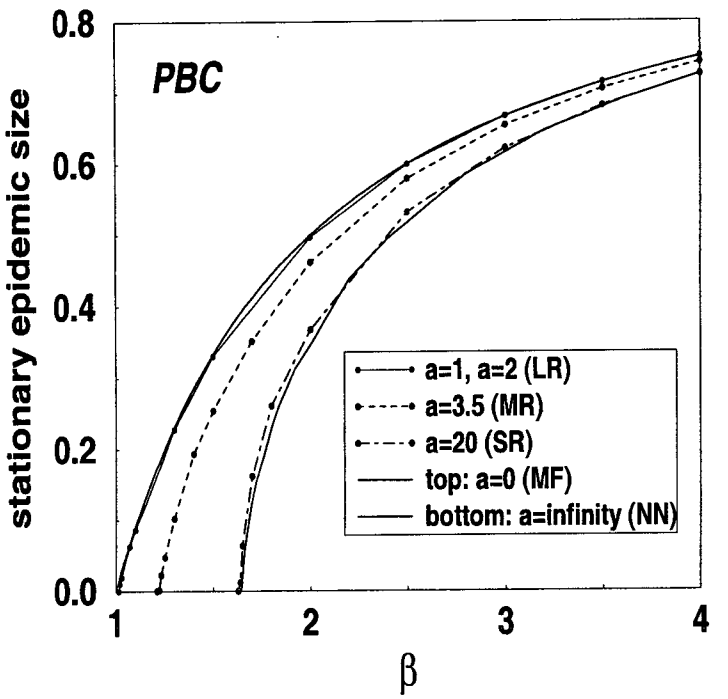


Figure 3.1: *Expected stationary epidemic size against infectiousness (phase diagrams). PBC.*

exponents $a = 1, 2, 3.5, 20$, in a PBC lattice of size $L = 200$; curves corresponding to $L = 100$ differ very little from these.

The results have the following features. Different interactions exhibit similar behaviour, with a threshold value of the infectiousness, β_c, separating regions of epidemic persistence and extinction. Quantitatively, epidemics are larger and the threshold smaller (Fig. 3.2) when the interaction range is larger. This is caused by a reduction in the aggregation of infectives with decreasing a (see also the epidemic snapshots in Fig. 3.31), which increases the likelihood of pathogens being deposited on susceptibles rather than on infecteds. All curves in Fig. 3.1 are bounded below by the NN model ($a = \infty$), which is closely approached by the $a = 20$ model, and above by the MF model ($a = 0$). Curves for $a = 1$ and

$a = 2$ are indistinguishable and nearly coincide with the MF model, in agreement with the result (2.34) proven in chapter 2 which states that in the limit $L \to \infty$ the behaviour of μ is the same as for MF $\forall\ a \leq 2$.

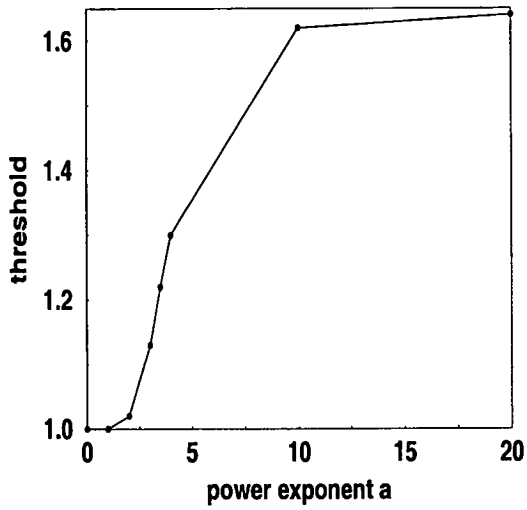


Figure 3.2: *Infectiousness threshold for persistence against power exponent a.*

The steepness of the curves near the persistence threshold reveals great sensitivity to variation or uncertainty in β. Sharp, infinite–slope transitions are typical of epidemic models with NN interactions (Levin & Durrett, 1997; Filipe & Gibson, 1998) and of critical phenomena in physical systems (Stanley, 1972). Fig. 3.1 shows that they are a feature of more general interactions. The asymptotic result mentioned above implies that, in the limit $L \to \infty$, the slope of the curves $(\partial\mu/\partial\beta)$ at the critical threshold is 1 for $a \leq 2$ and infinite otherwise. This divergence relates to the emergence of long–range correlations near the threshold.

In the previous chapter, the contact distribution (2.3) was defined as a probability distribution normalised over all possible deposition sites in a domain D. The normalisation factor $Z(a, L)$ could be absorbed into parameter β, but there are reasons for not doing so. First, only with Z is the model mathematically well–defined in the limit $L \to \infty$ for $a \leq 2$. Second, we wish to compare behaviour corresponding to different values of L and a. Let us suppose that no explicit normalisation Z had been used in (2.3) and that $\beta' = \beta/Z(a, L)$ is the corresponding infection rate. This approach has been used for studying other models with long–range interactions (Cannas & Tamarit, 1996). Since β' depends on a and L, it cannot be interpreted as the rate of propagule production. Fig. 3.3 displays the same curves as Fig. 3.1, but now the stationary epidemic size is plotted against β'. The figure shows that the new phase diagrams have undesirable features:

1. much greater sensitivity to variation in L and a;

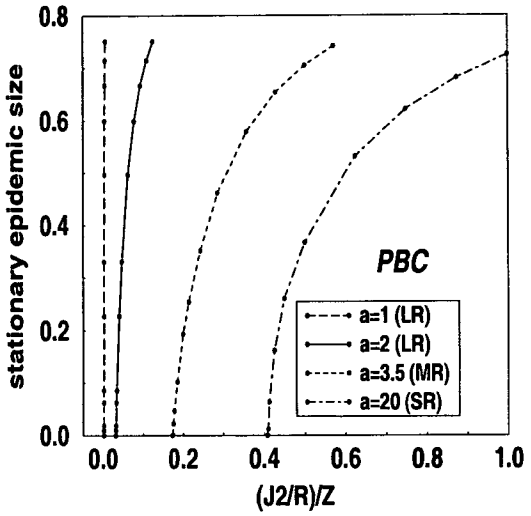


Figure 3.3: *Stationary epidemic size against $\beta' = \beta/Z$. PBC, L=200.*

2. for $a \leq 2$, in the limit $L \to \infty$ (when $Z \to \infty$), they collapse onto a single vertical line at $\beta' = 0$ and all the information about behaviour in this region is lost.

3.5.2 Boundary effects on epidemic size

We now consider a lattice with FBC. Two factors contribute to smaller probabilities of infection in this case. First, some pathogen propagules are deposited outside the boundary: pathogens produced at $\mathbf{x}$ are dispersed over an area D, but only a fraction $\rho(\mathbf{x})$ is deposited inside the system ($\rho(\mathbf{x}) = 1$ with PBC); Fig. 3.4a depicts how $\rho(\mathbf{x})$ varies along a diagonal cross–section of the lattice. Second, the system is no longer homogeneous: at the lattice centre, interactions with all other sites are much the same as in a PBC lattice, but sites nearer the boundary are less exposed and thus less likely to become infected. For example, the maximum infective challenge $\Phi_{\mathbf{x}}^{max} = J_2 \sum_{\mathbf{y} \neq \mathbf{x}} f(|\mathbf{x} - \mathbf{y}|)$ and the density of infecteds vary considerably across the system (Fig. 3.4b), whilst in a PBC lattice there is no overall variation. Fig. 3.4c shows the corresponding snapshot.

Fig. 3.5a compares μ for PBC and FBC lattices (with $L = 200$) in LR and SR dispersal cases. In the LR case, the presence of boundaries significantly reduces epidemics and increases the threshold β_c. The same comparison with $L = 100$ (not shown) yields very similar curves but with slightly greater differences between PBC and FBC lattices, i.e. edge effects decrease with system size. Fig. 3.5b compares μ in FBC lattices of different sizes: epidemics are marginally larger in the larger system. The corresponding curves in PBC lattices (not shown) are

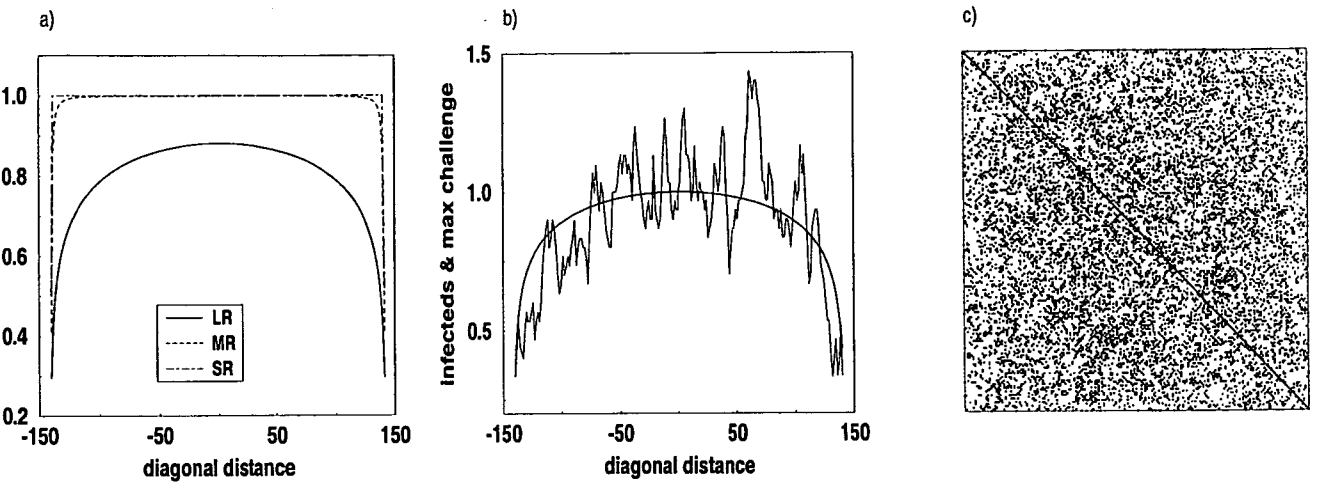


Figure 3.4: *FBC. a) Fraction of propagules produced at* $\mathbf{x}$ *which are deposited in the system: variation with distance from lattice centre along a diagonal cross-section; b) Density of infecteds in a* 10×10 *quadrat and maximum infective challenge* $\Phi_{\mathbf{x}}^{max}$ *along the same diagonal (divided by value at the lattice centre; c) Disease snapshot corresponding to b).*

almost identical, i.e. system–size effects are reduced with PBC. In conclusion, edge effects (Fig. 3.5a) are more important than size effects (Fig. 3.5b), both being more noticeable with LR than SR dispersal. With MR dispersal the effects are intermediate.

Fig. 3.6 is the same as Fig. 3.1 for an FBC lattice. An important difference relative to the PBC lattice (also apparent in Fig. 3.5a) is that, for given β, the epidemic level does not grow monotonically with increasing dispersal range (decreasing a). In this case, curves cross: $a = 2$ and $a = 20$ at $\beta \simeq 2$, and $a = 2.5$ and $a = 3$ at $\beta \simeq 1.3$ (not shown). Also, $a = 2$ is below $a = 3.5$ right from the threshold, which is the reverse of Fig. 3.1. The reason for this behaviour is that, for any given rate of pathogen production β, a fraction of the propagules is now lost to the outside and does not contribute to infections. This loss increases with increasing dispersal range (especially for $a < 2$) and with increasing β.

3.5.3 Variability in epidemic size

Population behaviour exhibits increasing variability with increasing proximity to extinction or decreasing dispersal range (i.e. with increasing spatial aggregation). This is illustrated in Fig. 3.7, which shows the variance against the mean of the epidemic size, μ (each point was estimated, from the long–term behaviour of hundreds of realisations, as in the transient behaviour). In the SR dispersal case, the variance is larger and it increases faster with decreasing disease level than in

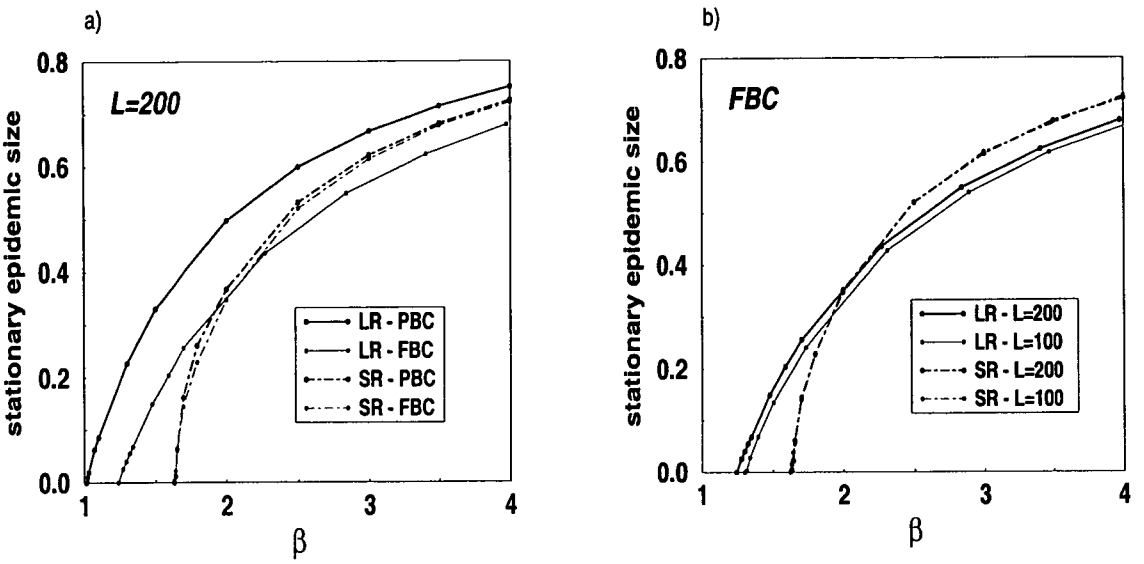


Figure 3.5: *Expected stationary epidemic size against infectiousness. a) edge effects; b) size effects.*

the LR and MR cases. For LR dispersal, the relation between the variance and mean of the epidemic size is almost linear.

3.5.4 Spatial correlations

We use the spatial autocorrelation function $C(r)$ (see (2.19) in section 2.4) to analyse the distribution of disease in the stationary regime. The nature and dynamics of spatial correlations and their effect on disease progress is a fascinating though not well understood problem which has received increasing attention in plant epidemiology in recent years (Reynolds & Madden, 1988; Xu & Ridout, 1996; Hughes *et al.*, 1997; Ferrandino, 1996, 1998; Filipe *et al.*, 2000). We consider lattices of size $L = 200$ with $\mu = 4\%, 10\%$ and 25% in each of the three dispersal cases. Fig. 3.31 shows corresponding patterns of disease. Fig. 3.8 depicts typical raw correlation data and the corresponding correlations smoothed through the use of histograms (see (2.7) and end of section 2.3). Hereafter only smoothed data are shown.

We distinguish between lattices with (FBC) and without (PBC) boundaries: spatial correlations are estimated using definitions (2.17) and (2.18) in the first case and (2.19) in the second case.

3.5.4.1 Spatial correlations in unbounded systems

We first consider the homogeneous PBC lattice. Figs 3.9 and 3.10 compare $C(r)$ for given epidemic levels and given interactions, respectively. The following con-

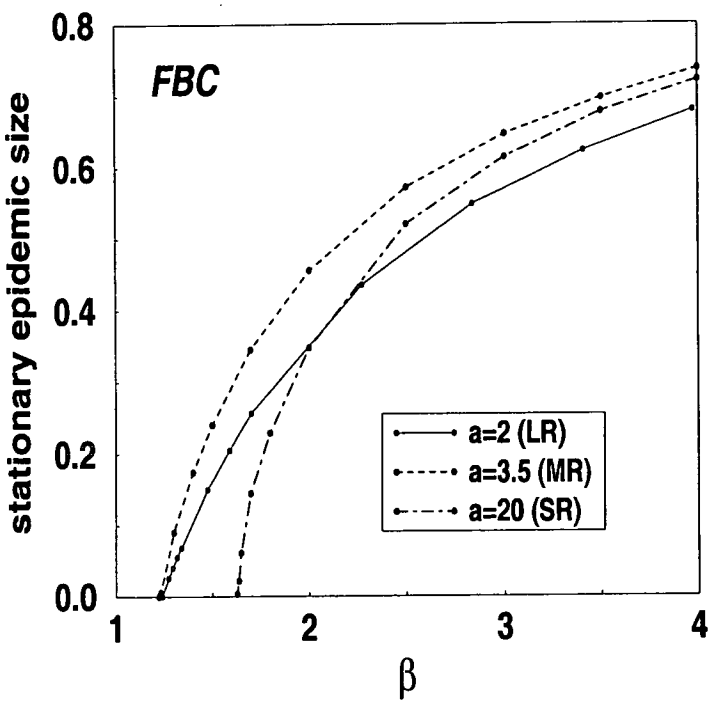


Figure 3.6: *Expected stationary epidemic size against infectiousness (phase diagrams). FBC.*

clusions can be drawn and related to previous results.

- Correlations, especially at short distances, increase with decreasing range of dispersal (larger a). There is a corresponding increase in local aggregation (Fig. 3.31) and in the strength of infection β needed to reach a given epidemic level (Fig. 3.1). The variation of $C(1)$, which may be regarded as an index of aggregation between infecteds, is sketched in Table 3.1.

- Correlations, especially at large distances, increase as extinction is approached. A corresponding emergence of structure at large scale is observed (Fig. 3.31). The high sensitivity of μ to variation in β near extinction (Fig. 3.1) is thus explained by the appearance of long–range correlations which amplify local fluctuations in disease. Such large–scale behaviour is typical of critical phenomena at continuous phase transitions (Stanley, 1972).

- With LR dispersal, correlations are very small and vary little with the epidemic level. This agrees with Figs 3.31 and 3.1 which suggest that infecteds and susceptibles are well mixed in this case.

A more quantitative analysis follows from studying the functional form of the correlations, e.g. exponential or power–law. With LR dispersal ($a = 2$), ln–ln plots of $C(r)$ are nearly linear for distances up to 20 for all values of μ (Fig. 3.11).

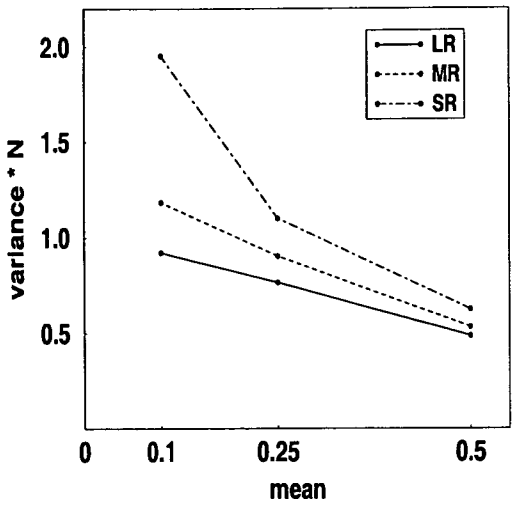


Figure 3.7: *Variance (times the number of individuals N) against mean of epidemic size, μ. Samples of 128 or 256 realisations with PBC and $L = 100$.*

$C(1)$	$a = 2$	$a = 3.5$	$a = 20$
4%	0.033	0.209	0.369
10%	0.031	0.193	0.338
25%	0.026	0.147	0.270

Table 3.1: *Measure of aggregation: correlation at lag* 1, $C(1)$.

In order to introduce a length scale, ξ, we fit the data with the ln–ln transform of the function:

$$\tilde{C}(r) = \frac{C(1)}{[1 + (r-1)/\xi]^b}. \tag{3.3}$$

Similar results are obtained fitting a straight line ($\xi = 1$) or allowing ξ to vary, in which case $\xi \simeq 1.1$ for all values of μ. Exponent b increases, i.e. the range of correlations decreases, with increasing μ (Table 3.2; values are similar with either $\xi = 1.1$ or 1). Fig. 3.12 compares correlations for $a = 2$ and $a = 1$ and shows that there are effectively no correlations with $a = 1$ ($C(1) = 10^{-3}$). These results tell us something unexpected: while the epidemic size is indistinguishable for all $a \leq 2$ and behaves as if the disease distribution were random (see Fig. 3.1 and section 2.5.2), the correlation analysis suggests a substantial difference between disease patterns with $a = 2$ and $a = 1$.

Qualitatively different correlations are found with MR and SR dispersal. Fig. 3.13 displays the corresponding ln–linear plots, which are very well fitted by

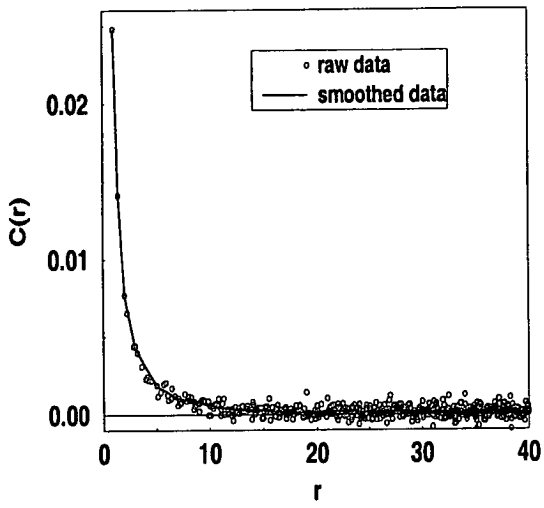


Figure 3.8: *Spatial correlations in the stationary regime for LR dispersal and epidemic size* 25%. *Raw and smoothed data.*

b	$a = 2$	$a = 3.5$	$a = 20$
4%	1.55	0.51	0.59
10%	1.59	0.66	0.62
25%	1.81	0.67	0.86

Table 3.2: *Best fit values of exponent b in (3.3) and (3.4).*

stretched exponentials,

$$\tilde{C}(r) = C(1)\, e^{-\left(\frac{r-1}{\xi}\right)^b}, \tag{3.4}$$

where ξ is a correlation length scale. Power–laws and simple exponentials yield poorer fits to data (not shown). Fitted values of ξ range from 1.1 to 2.2. Fitted values of b are given in Table 3.2. As expected, curves approach simple exponentials ($b = 1$) and the correlation length decreases as epidemic size μ increases. Conversely, as μ tends to zero, a crossover to power–law correlations might be expected, which is consistent with sequences of ln–ln plots of $C(r)$ down to $\mu = 4\%$. Fig. 3.14 illustrates this point in the SR dispersal case: ln–ln plots of correlations are nearly linear for $\mu = 4\%$. Sampling behaviour near extinction, however, is very difficult because correlations in SR and MR interacting systems develop oscillations which reflect the presence of structure at scales comparable to L (Fig. 3.31).

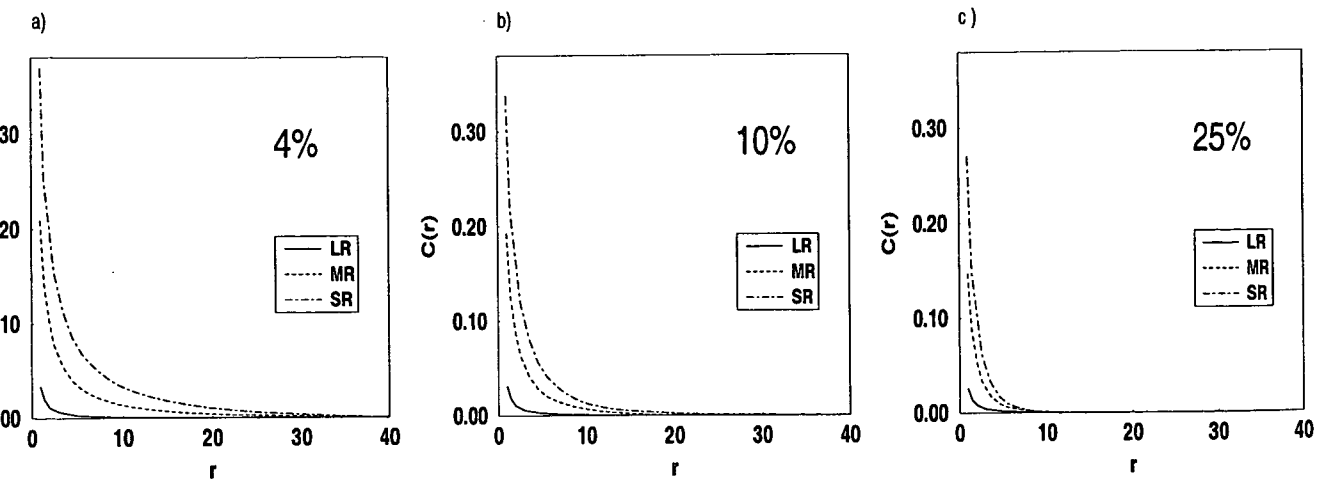


Figure 3.9: *Variation of spatial correlations in the stationary regime with the interaction (SR, MR and LR), for given epidemic levels. a) μ = 4%; b) μ = 10%; c) μ = 25%.*

3.5.4.2 Spatial correlations in bounded systems

We now consider a heterogeneous FBC lattice. In this case the density of disease patterns is heterogeneous, as demonstrated by Fig. 3.4b which shows the local density of infecteds decreasing as the boundary is approached. Naturally, this effect is strong with LR dispersal but negligible with SR or NN dispersal (Filipe & Gibson, 1998). The autocorrelation $C(r)$ quantifies correlations between local variables relative to their local mean (over realisations). Since the lattice is non–uniform, the local mean is also non–uniform, which is why (2.18) is required for estimating the expectation of $C(r)$. Using (2.19) to estimate correlations on a lattice with FBC leads to erroneous results, especially for LR dispersal, as illustrated in Fig. 3.15.

Fig. 3.16 compares $C(r)$ in PBC and FBC lattices, with $\mu = 25\%$, estimated using (2.19) and (2.18) respectively. Apart from a greater residual dispersion in the FBC case, the two raw–data sets are very similar. Hence, for these (and possibly most) parameter values there is little difference in the correlation of disease patterns of bounded and periodic systems with the same fraction of infecteds. The latter condition is essential for a valid comparison since correlations depend on pattern density. In fact, by comparing systems with the same stationary level of disease, some of the boundary effects on the epidemic process are accounted for, so correlations should not be much affected provided their range is smaller than L. We recall (see Fig. 3.5) that the same epidemic size is reached for a larger value of β with FBC than with PBC.

In practice, it is often necessary to analyse a single pattern produced by a

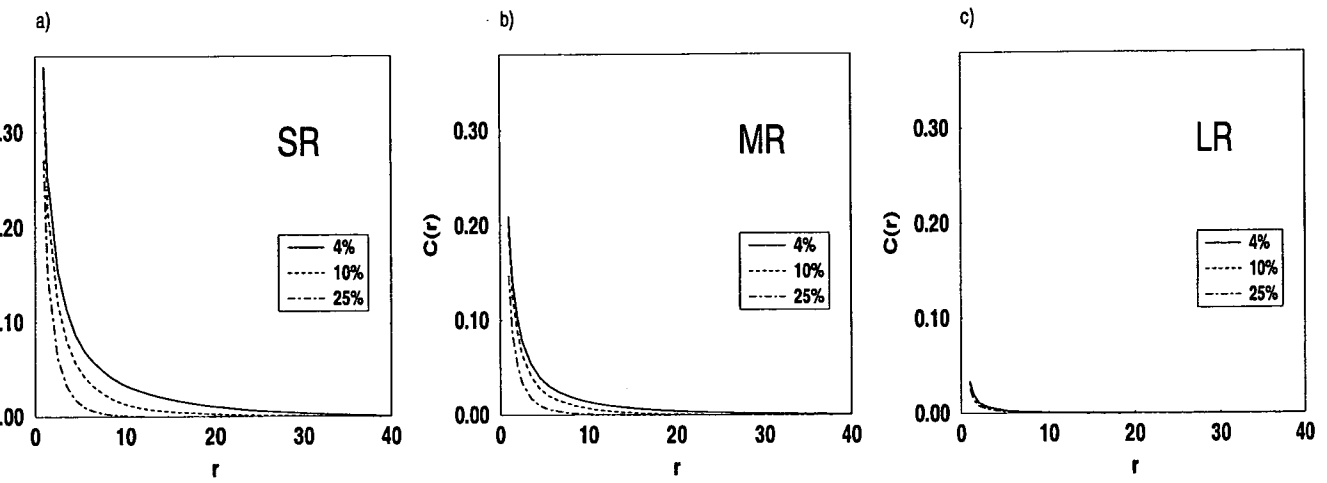


Figure 3.10: *Variation of spatial correlations in the stationary regime with the epidemic level ($\mu = 4\%, 10\%, 25\%$), for given interactions. Note that in the LR case correlations are very small and do not vary much with μ. a) SR; b) MR; c) LR.*

stochastic process. Equation (2.19) can be applied to a single pattern provided that the expectations over realisations (and the average over time) are removed, giving $C(r) = (\rho_{II}(r) - \rho_I^2)/(\rho_I - \rho_I^2)$, where $\rho_{II}(r)$ and ρ_I are the spatial averages defined by (2.12) and (2.13). In contrast, the averages involved in (2.18) can only be evaluated when a set of realisations is available. In the latter case, other estimators of correlations might be applied. Examples can be found in Filipe *et al.* (2000).

3.5.5 Theoretical prediction for the critical threshold

In the previous chapter we used the equations for the time evolution of the average fraction of infecteds (2.30) to predict the critical threshold for epidemic persistence, $\beta_c = 1/(1 - \Gamma_c)$, where Γ_c is defined by (2.31) and evaluated at the threshold. The stationary epidemic size is given by:

$$\mu = \begin{cases} 1 - \dfrac{1}{\beta\ (1 - \Gamma)} & \text{if } \beta > \beta_c \\ 0 & \text{if } \beta \leq \beta_c. \end{cases} \tag{3.5}$$

We recall that these results are valid for unbounded homogeneous lattices.

We tested this simple model by comparing theoretical predictions for μ against simulated values. To estimate Γ we substituted in (2.31) the corresponding simulated spatial correlations $C(r)$. Fig. 3.17 shows remarkable agreement, supporting the validity of relationship (3.5) between infection level and spatial correlations.

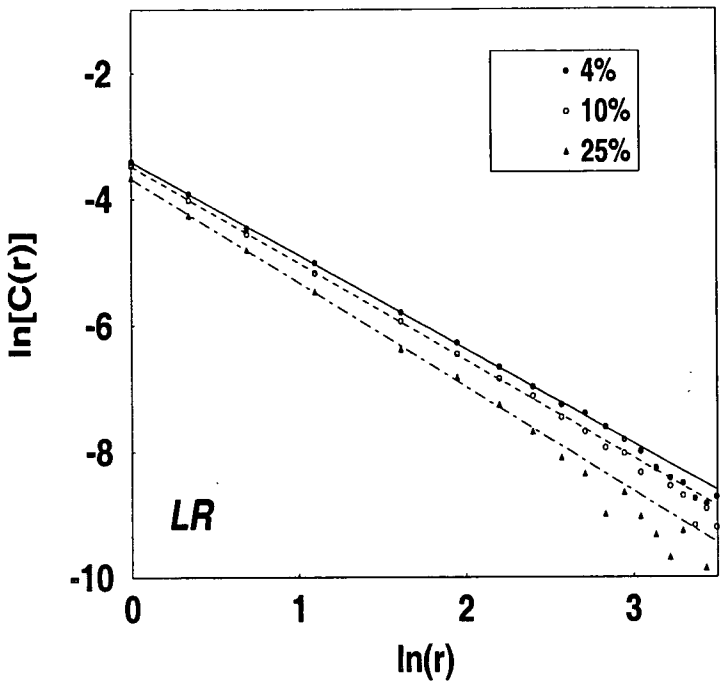


Figure 3.11: *Ln–ln plots of $C(r)$ for LR dispersal ($a = 2$). Linear fits have slope b given in Table 3.2.*

3.5.6 Patterns of disease

Fig. 3.31 shows disease patterns in the stationary regime simulated on a 200×200 square lattice. Each pattern refers to one of the three dispersal cases and one of the three epidemic sizes ($\mu = 4\%$, 10% and 25%) corresponding to the spatial correlations in Figs 3.9 and 3.10.

Different forms of dispersal produce very different patterns of disease. The effectiveness of a dispersal mechanism in spreading disease (i.e. in producing infective–susceptible interactions) is in turn influenced by the type of spatial pattern which it generates. In other words, disease progress is conditioned by its distribution.

LR dispersal produces patterns of disease in which the distribution of infecteds looks almost random. This is in good agreement with Fig. 3.1, which exhibits similarity between the MF model and the LR dispersal model ($a = 2$). It is also in good agreement with Figs 3.9 and 3.10, which show that spatial correlations are very small for LR dispersal, although they are not zero, as revealed by Fig. 3.12. As the range of interaction decreases, aggregation increases. SR dispersal generates the most clustered distributions of infecteds, which are the cause of the large spatial correlations observed in Figs 3.9 and 3.10.

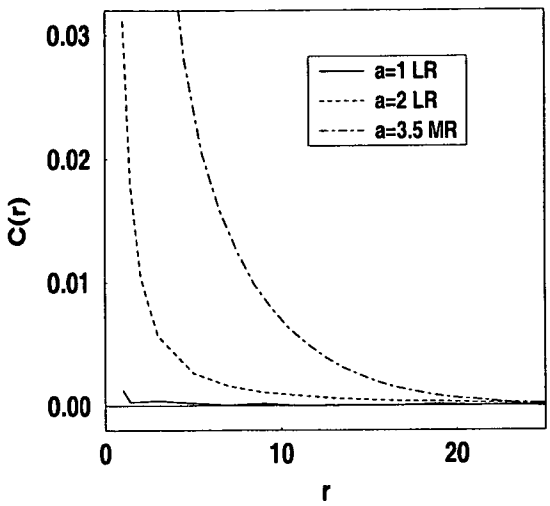


Figure 3.12: *Correlation functions for* $a = 1$, $a = 2$ *and* $a = 3.5$. $\mu = 10\%$.

3.6 Transient behaviour

We now analyse the transient behaviour. First we consider the SIS model which includes recovery of infecteds. Epidemics are initiated with 20 randomly distributed infecteds on a lattice of linear size $L = 100$ and, as before, we study the epidemic size and spatial correlations. We then investigate the effects of different sets of initial conditions, and finally we analyse the SI model (no recovery). In the following, only unbounded lattices (PBC) are considered. Since transients are characterised by wide stochastic variability (Fig. 3.18), we sample behaviour using large numbers of stochastic realisations (from 128 to 256); sampling is conditional on epidemic survival.

3.6.1 Epidemic size

First we look at the distribution of epidemic sizes over the sample of simulated stochastic realisations, at three points in time. Times are chosen according to the shape of the standard deviation $\sigma(t)$ (Fig. 3.20), and correspond to the half–peak (t_1) and peak (t_2) of σ and to a point in the stationary regime (t_∞). Fig. 3.19 displays the evolving distribution, after appropriate rescaling of the axes, in the case of LR dispersal and 50% long–term infection. Equilibrium distributions are nearly normal, as confirmed by logarithmic plots (not shown). The roughness of the transient distributions suggests that sample size might not be large enough, so we cannot conclude whether the distribution is normal, although we see no reason in principle why it should not be. The distribution with the lowest average (time t_1) appears to be skewed towards positive values, which might be explained by

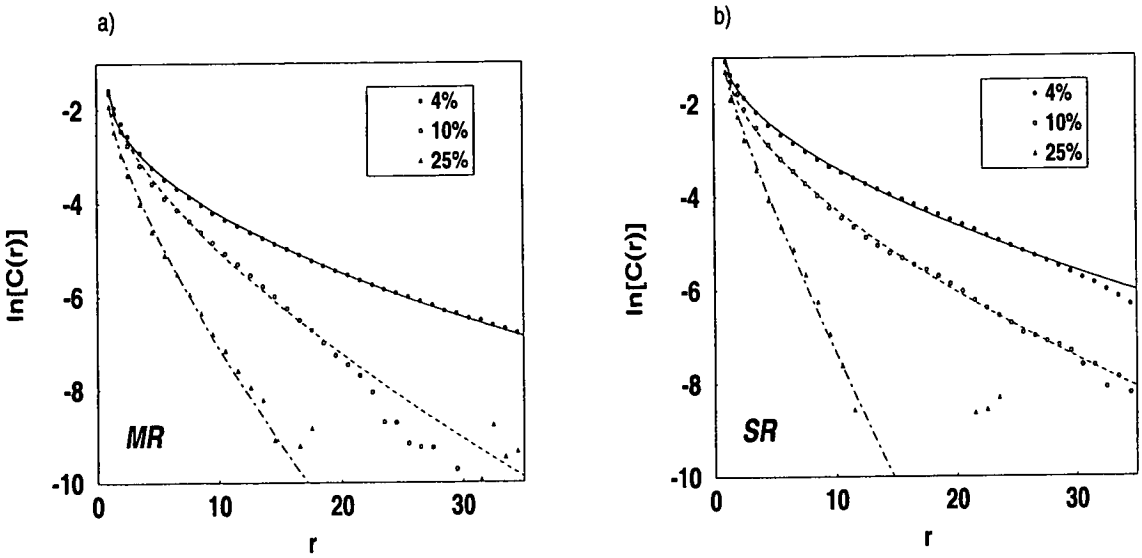


Figure 3.13: *Ln–lin plots of $C(r)$ for: a) MR, and b) SR dispersal. Lines are stretched exponentials fits (see (3.4) and Table 3.2).*

the increase in the probability of extinction when the epidemic size is small.

Next, we consider the first moment of the above distribution. Fig. 3.29 shows the *mean* of the epidemic size $\mu(t) \equiv \langle \rho_I(t) \rangle$ for LR, MR and SR dispersal and long–term epidemic sizes 10%, 25% and 50%. Also shown are: a few error bars $\pm\, \sigma[\rho_I(t)]$, the corresponding Mean Field (MF) predictions, and least square fits of the following generalised (asymmetric) logistic function, also known as Richards curve (Campbell & Madden, 1990),

$$\mu(t) = \frac{\mu_\infty}{[1 + [(\mu_\infty/\mu_0)^{1/\alpha} - 1]\; e^{-t/\tau}]^\alpha}, \tag{3.6}$$

with fitting parameters α and τ, and inflection point $\mu_{il} = \mu_\infty[\alpha/(1+\alpha)]^\alpha$. MF predictions, which are common in epidemiology (e.g. Mollison, 1995), are based on the assumption that there are no spatial correlations and thus on expressing the growth rate of the mean number of infecteds as proportional to the mean number of infecteds multiplied by the mean number of susceptibles (for a more detailed explanation of the MF approximation see chapter 4). This assumption appears reasonable in the case of LR dispersal but substantially overestimates the long–term epidemic size (Filipe *et al.*, 2000) and the rate of epidemic growth in the MR and SR dispersal cases. The logistic fit is also good with LR dispersal but is much less satisfactory with MR and SR dispersal. The reason for this is related to the reason for the failure of the MF prediction. Equation (3.6) is in

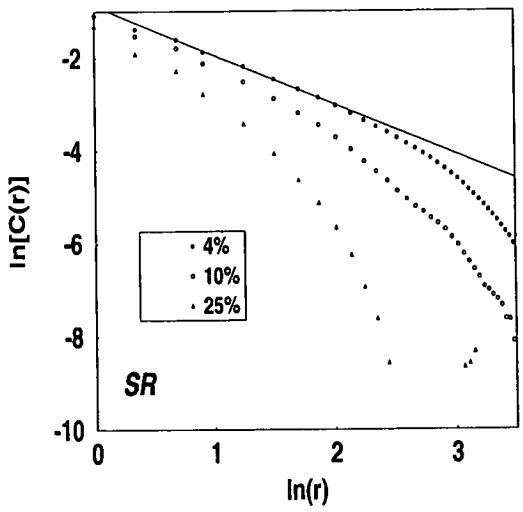


Figure 3.14: *Ln–ln plots of $C(r)$ for SR dispersal ($a = 20$). Correlations tend to decay as power–laws as μ tends to 0. The linear fit for $\mu = 4\%$ has slope $b = 1.06$. Larger epidemic sizes correspond to correlations which decay more like exponentials.*

fact the solution to the generalised MF equation:

$$\frac{d\mu(t)}{dt} = \frac{\alpha}{\tau}\,\mu(t)\left[1 - \left(\frac{\mu(t)}{\mu_\infty}\right)^{1/\alpha}\right]. \tag{3.7}$$

The main features of $\mu(t)$ are its long–term limit, studied in the previous section, and the time scale of the transient. We shall try to define a time scale (average "transient duration") below.

We now concentrate on the *standard deviation* of the epidemic size, $\sigma(t)$. Fig. 3.20 shows σ against time in the nine cases of dispersal vs long–term infection. The main features of $\sigma(t)$ are a well–defined peak (t_p, σ_p) during the transient, and the long–term value σ_∞. For clarity, axes were rescaled as σ/σ_∞ and t/t_p. Note that equilibrium is typically reached by $t = 2.5\,t_p$. The curve features σ_p, σ_∞ and t_p vary with the epidemic parameters a and β (i.e. μ_∞), as shown in Fig. 3.21.

3.6.2 Transient duration

From Fig. 3.29 it is clear that each mean curve has its own time scale and shape. We would like to define a transient duration in order to study its dependence on the basic parameters and to establish qualitative differences in the shape of disease progress curves. Obvious definitions of a time scale invoke either the inflection or the half–height point of $\mu(t)$, or the same features of its logistic fit. We find, however, that none of these curve–descriptive approaches appear to

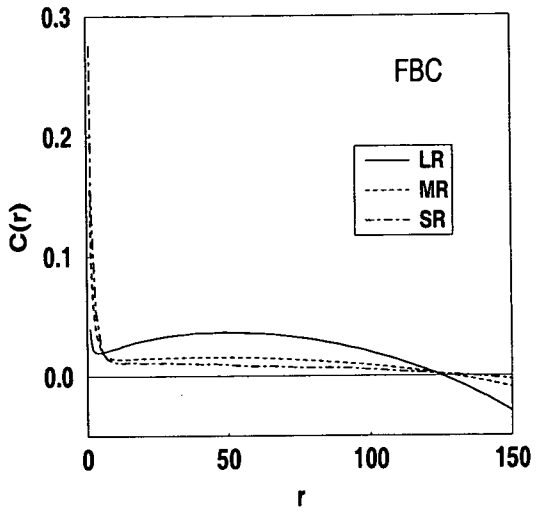


Figure 3.15: *Correlations estimated with (2.19) on a bounded lattice for the three dispersal cases.*

capture the variation of the time scale with the model parameters. Here we take a conceptually different approach by exploiting the observed peak in $\sigma(t)$. This peak is a striking feature which can be measured more accurately than inflection points. Table 3.3 compares possible time scales associated with the curves $\mu(t)$: the peak of $\sigma(t)$, t_p, the estimated inflection point of $\mu(t)$, t_i, and the inflection point of the fitted logistic, t_{il}. We find that $t_{il} < t_i < t_p$. Times t_i and t_{il} are always very close, while t_i and t_p are close in the first five cases but differ considerably in the others where spatial correlations are known to be important. (The half-height time $t_{0.5} : \mu(t_{0.5}) = \mu_\infty/2$ does slightly better than t_i but also lags behind t_p.) This reinforces the idea that t_p captures aspects of the behaviour that t_i does not.

a	μ_∞	t_{il} (logistic)	t_i (μ)	t_p
2	50%	5.7	5.7	6.1
2	25%	15.0	15.1	16.3
2	10%	34.2	35.0	39.0
3.5	50%	9.0	9.4	10.3
3.5	25%	28.3	30.0	33.2
3.5	10%	68.2	73.0	85.6
20	50%	21.7	24.0	34.7
20	25%	71.6	82.0	110.8
20	10%	216.8	229.0	360.4

Table 3.3: *Measures of transient timescale.*

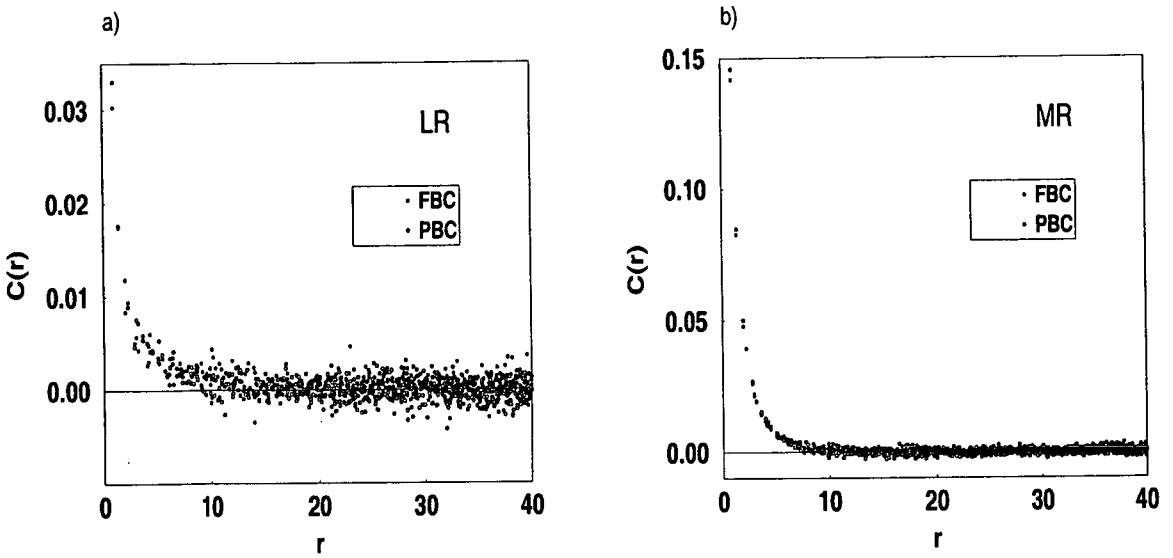


Figure 3.16: *Expected correlations in the stationary regime estimated with (2.19) with PBC and (2.18) with FBC ($\mu = 25\%$ for both lattices). a) LR; b) MR.*

For these reasons, we propose to use the peak of $\sigma(t)$ for defining transient duration. We also observed that there is a variable degree of asymmetry of $\sigma(t)$ around its peak, i.e. the time interval between the peak and some definition of the start of the equilibrium plateau varies with a and β differently than the interval between the start of the epidemic and the peak. This probably means that in the second stage of the transient the system has reached spatial configurations with different properties (i.e. closer to the average long–term configuration) and that the spatial processes have a different effect on epidemic progress. It also suggests that t_p does not contain enough information for defining transient duration.

In order to estimate the duration of the post–peak transient we consider $t_p + t_h$, the time at which σ reaches a certain value between σ_p and σ_∞. Specifically, we consider t_p (Fig. 3.22) such that:

$$\sigma(t_p + t_h) = \frac{\sigma_p + \sigma_\infty}{2}. \tag{3.8}$$

We then approximate the post–peak relaxation time of $\sigma(t)$ by $2t_h$, which is probably a slight underestimation. Finally, we define average "transient duration" as:

$$t_{tr} = A\ [t_p + 2t_h], \tag{3.9}$$

where $A = 1.5$. The precise value of factor A is somewhat arbitrary and not very important as it does not change the relative duration of different transients; the value chosen is suggested by examination of $\sigma(t)$ (Fig. 3.22). Fig. 3.23 shows the

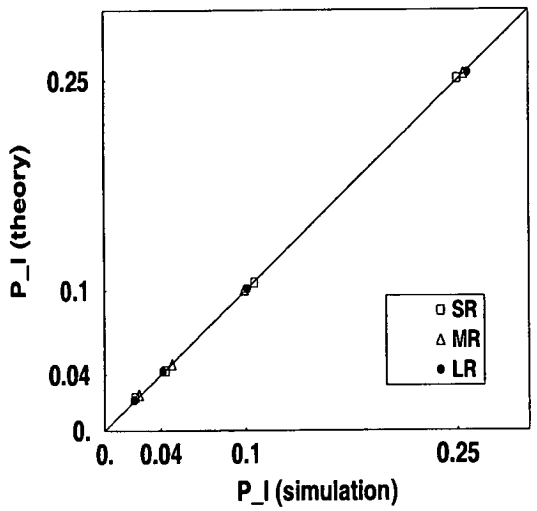


Figure 3.17: *Theoretical prediction for average stationary epidemic size μ (3.5), against values from simulation on an unbounded $L = 200$ lattice. The line (slope 1) is for reference.*

dependence of t_{tr} on the epidemic parameters. As expected, transients are longer when the dispersal range is shorter. They are also longer when the long–term epidemic size is smaller, which might sound counter–intuitive given the initial state ($P_I(0) = 0.2\%$). An explanation is that when the stationary epidemic size is smaller, long–term correlations are larger (see Figs 3.9 and 3.10), and consequently the stochastic build–up of such an attractor state takes longer than the build–up of a comparatively more random state.

Our final objective is to use our definition of transient duration for comparing disease progress curves with very different time scales. Figs 3.24a and b display some of the curves in Fig. 3.29 with axes rescaled in the form $[\mu(t) - \mu(0)]/[\mu_\infty - \mu(0)]$ and t/t_{tr}, for given dispersal function (MR) and long–term epidemic ($\mu_\infty = 25\%$), respectively. Sets of curves corresponding to other parameter values have similar appearances. Apart from their effect on μ_∞, dispersal processes can be distinguished by the rate of growth of μ (Fig. 3.24b): curves are more linear (and thus t_{tr} is larger) when the dispersal range is shorter. In the SR case, spread is dominated by peripheral growth of infected clusters within which the stationary density is rapidly reached (see Fig. 3.30), hence the growth rate is proportional to the perimeter rather than the area of clusters. In the LR case, there is a chain reaction yielding initial exponential growth and a sigmoid shape. The qualitative difference between curves with different long–term epidemic size (i.e. different β) and given a (Fig. 3.24a) is analogous; as explained above, growth is slower (t_{tr} is larger) when μ_∞ is smaller.

The fact that all curves closely merge as they approach stationarity (rather

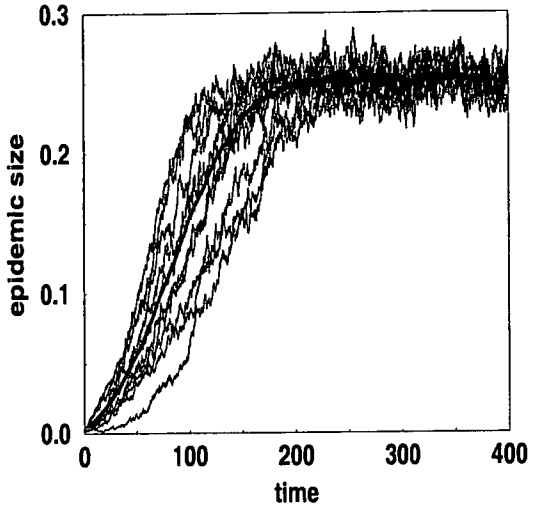


Figure 3.18: *Stochastically identical disease progress curves and their mean (bold curve). Disease spreads on a* 100×100 *square lattice according to a contact distribution with exponent* $a = 20$*, from an initial random distribution of* 20 *infectives. The mean long-term level is* 25%.

than saturating at different values of t/t_{tr}) is in our view an indication that the essential dependence of t_{tr} on the model parameters is being well captured by (3.9). Approaches based on the inflection or half–height point of $\mu(t)$ invariably lead to curve crossing in the middle region, another indication of how poorly they capture the parameter dependence of t_{tr}. Variations on definition (3.9) based on different combinations of t_p and t_h, e.g. $t_{tr} = 3\ t_p$ or $t_{tr} = 2\ [t_p + t_h]$, also produced less satisfactory rescaled curves $\mu(t)$ since they account less effectively for the asymmetry of $\sigma(t)$ about its peak.

3.6.3 Spatial correlations

The spatial correlation function $C(r,t)$, as defined by (2.19), is used to characterise the spatial structure of disease. $C(r,t)$ was recorded in each of 128 (or 256) stochastic realisations with $\mu_\infty = 25\%$, at times when the epidemic reached 5%, 10%, 15% and 20% levels. The evolution up to the stationary regime of the average of correlations over runs for the three dispersal cases is shown in Fig. 3.25. For each distance lag, up to about $r = 20$, correlations are quite large at early stages and decrease monotonically in time towards long–term values (with the exception of the 5% curve in the SR case, which decreases faster and crosses the other curves, Fig. 3.25c). Disease patterns with any given disease level in the transient stage have larger correlations than disease patterns in the stationary stage with the same epidemic size. This behaviour reflects the rich dynamic

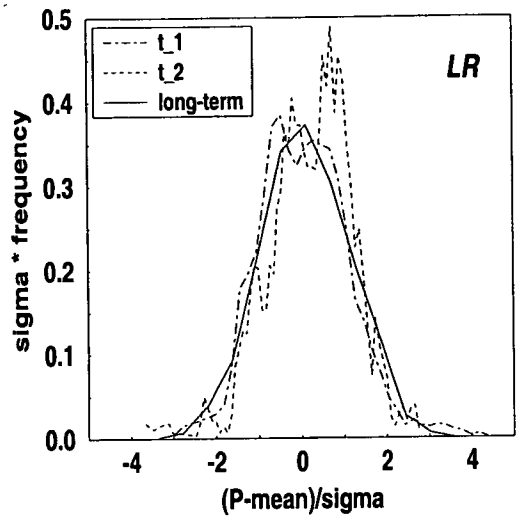


Figure 3.19: *LR. Relative frequency distribution of epidemic size values P_I at given times t_1, t_2 and t_∞. $\mu(t_1) = 0.08, \mu(t_2) = 0.26, \mu(t_\infty) = 0.50$.*

structure generated by the formation, growth and coalescence of secondary foci. Some of this complexity gradually fades out in time with increasing pattern density and random recovery. Complex dynamic structures were also reported in other studies (Shaw, 1995) in which a fractal (but not a correlation) analysis was carried out in the special case $a = 3$. Another notable feature of Fig. 3.25 is that all curves cross at about the same lag $r_c \simeq 30$ and are negative onwards. For these lags, anticorrelations decrease monotonically in time to zero, the long–term value for $r > 30$. This negative tail may be attributed to the heterogeneity of transient disease patterns (Fig. 3.30). Why all curves appear to cross about the same r_c is unclear.

Ln–ln plots corresponding to Fig. 3.25, for lags up to r_c ($C(r) > 0$), are shown in Fig. 3.26. Transient correlations are very nearly power–laws not only for LR dispersal, in which case long–term correlations also decay as power–laws (see Fig. 3.11), but also for MR and SR dispersal, where long–term correlations decay as stretched exponentials (see Fig. 3.13). This supports the above observation about the complexity of transient patterns and is related to their fractal properties (Shaw, 1995). This behaviour is, of course, bounded by the length scale r_c.

The above correlation functions are estimated means of the underlying distributions of $C(r,t)$ for each distance lag and time. We looked at the distribution of $C(r,t)$ over realisations at given times for given lags and found it to be fairly symmetric and to spread over positive values even when the mean is negative (not shown). One striking feature was that the distributions with zero mean (i.e. for $r = r_c$) had much sharper peaks than the distributions for other lags.

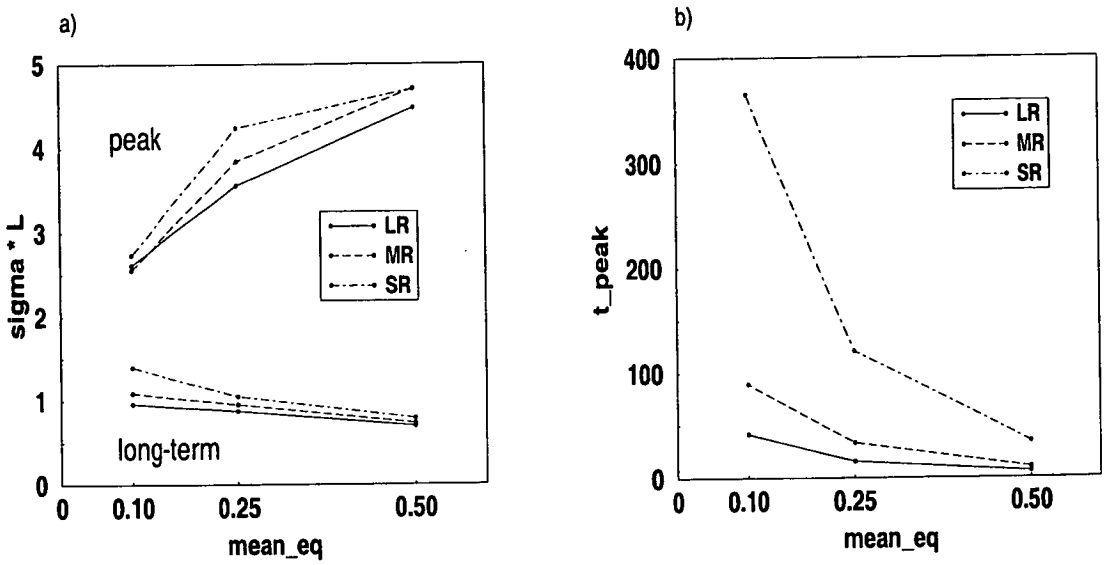


Figure 3.21: *Variation of the features of $\sigma(t)$ curves with epidemic parameters. a) Values at peak (σ_p) and long-term (σ_∞); b) peak time t_p.*

applicable in cases with recovery (because $\mu(t)$ fluctuates with finite sampling).

Variation in epidemic progress between realisations is at least as large as with recovery, as indicated by the values of σ_p (note that, apart from the long–term behaviour, $\mu(t)$ and $\sigma(t)$ resemble the case $\mu_\infty = 50\%$). Hence, although $\mu(t)$ is smooth, the absence of recovery does not make the process more deterministic.

3.6.5 Dependence on initial conditions

The initial number and distribution of infecteds are expected to affect transient behaviour (Xu & Ridout, 1998; Filipe & Maule, 2000 a). Results so far were obtained for a single set of initial conditions (IC20r) with 20 randomly distributed infecteds ($P_I(0) = 0.2\%$). Whilst an apparently natural initial condition would be a lattice with a single infected, beginning with 20 rather than 1 or 2 infecteds substantially decreases the probability of early extinction. It also decreases transient variability, making our samples more representative.

Here, we analyse the effects of having other initial conditions:

- 10 randomly distributed infecteds (IC10r);
- 2 randomly distributed infecteds (IC2r);
- 20 spatially correlated infecteds (IC20c), obtained by starting the process with IC2r and $R = 0$, and setting $R = 1$ and $t = 0$ when 20 hosts are infected.

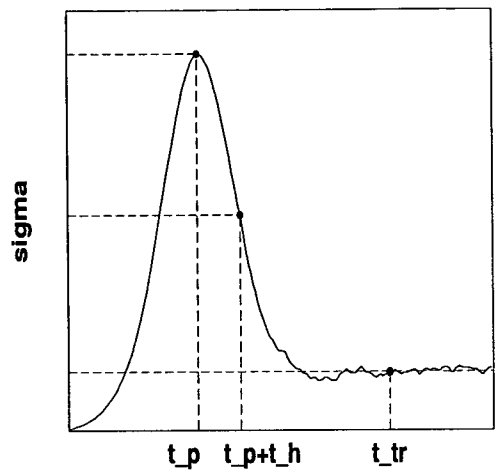


Figure 3.22: *Peak and post–peak half–height points of a generic standard deviation* $\sigma(t)$.

The epidemic died out quite often at early transient stages, especially when initiated with 2 infecteds only. All expectations shown are conditional on epidemic persistence. We illustrate the analysis in the extreme cases of LR and SR dispersal, with $\mu_\infty = 50\%$.

Fig. 3.28a compares the average disease progress $\mu(t)$ resulting from each initial condition. Relative to their respective time scales, the effects of the different initial conditions appear to be similar in the SR and LR cases: there is a delay in disease progress relative to IC20r which grows in the order: IC10r, IC2r. Effects of IC20c depend on the dispersal, which determines spatial correlations and thus the difference between IC20c and IC20r. In the LR case, IC20c and IC20r yield the same curve (indistinguishable in Fig. 3.28a) because initial correlations are very small (see above). In the SR case, the IC20c curve lies between the IC10r and IC2r curves: although the initial number of infecteds is larger than in IC10r, these are distributed in one or two clusters generated from the two initial infecteds; since new infections can only occur at the edge of the clusters, the epidemic is slower.

The standard deviation $\sigma(t)$ is qualitatively similar for different initial conditions (not shown). Its peak time t_p and width increase relative to IC20r in direct relation to the amount of transient delay in Fig. 3.28a. Applying definition (3.9) of transient duration, and rescaling $\mu(t)$ and t as in Fig. 3.24, yields the results in Fig. 3.28b for the SR case; the curves for IC20r and IC10r coincide. In the LR case (not shown), all rescaled curves nearly coincide (except for IC2r which, as in Fig. 3.28a, has slower growth). Therefore, definition (3.9) of t_{tr} also seems to work well for different initial conditions. The magnitude of $\sigma(t)$, i.e. σ_p, was

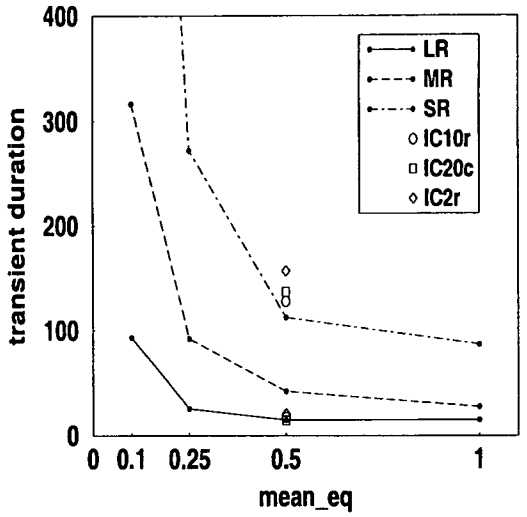


Figure 3.23: *Dependence of transient duration (estimated with (3.9)) on model parameters. Curves with dark dots correspond to the current case with $R = 1$ and initial condition IC20r (see section 3.6.5); dark dots at $\mu_\infty = 1$ correspond to the model with $R = 0$; open symbols correspond to the model with $R = 1$ but different initial conditions (IC10r, IC20c, IC2r).*

smallest in the case IC20r and significantly larger in the case IC2r, which was one of the reasons for choosing IC20r in the general study.

3.7 Conclusions

We have used simple stochastic spatial models to analyse the effects of a range of power–law contact distributions and sporulation rates on temporal and spatial aspects of plant epidemic development. We have investigated both the stationary behaviour (in the presence of a control process) and the transient behaviour (with and without a control process). Our main findings and conclusions are made below.

Stationary behaviour. In periodic systems, the mean epidemic size μ_∞ increases monotonically with β and a (which parameterise the infectiousness and dispersal range, respectively) in the way expected. The curves $\mu_\infty(\beta)$ vary continuously with a between the nearest neighbour and MF limits. The curves become indistinguishable for all $a \leq 2$, in agreement with (2.34), which states that for $a \leq 2$ μ_∞ converges to the MF prediction in the limit of an infinite system. In other words, the epidemic size of large systems is accurately predicted by the MF model when the dispersal process is sufficiently long–ranged (Fig. 3.1).

In systems with boundaries, epidemic size is smaller (and the threshold β_c is

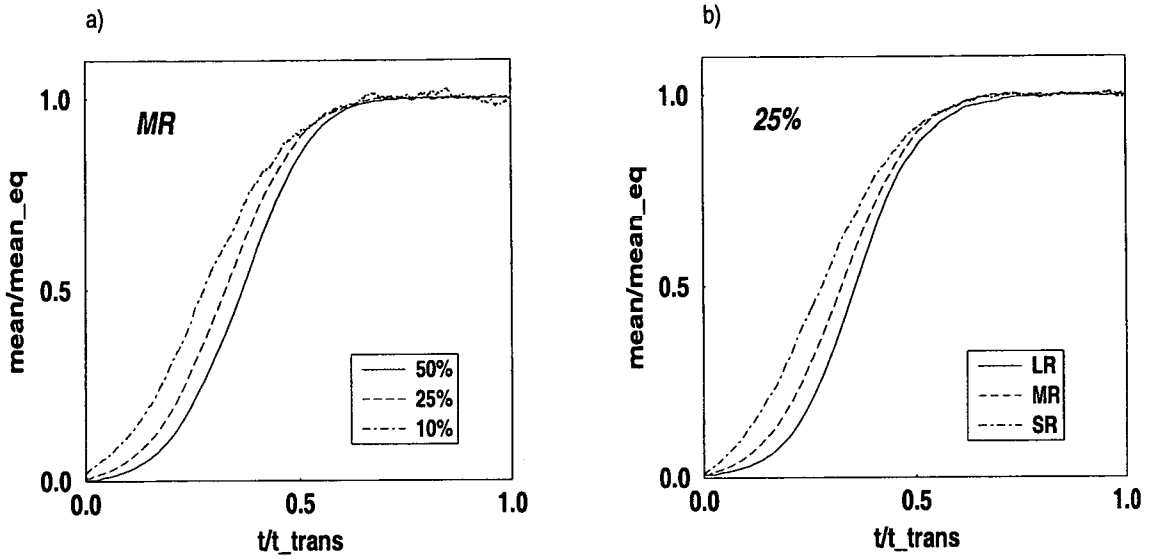


Figure 3.24: *Rescaled mean disease progress curves. a) MR; b)* $\mu = 25\%$.

larger) than in periodic systems. In addition, μ no longer varies monotonically with a because some propagules are dispersed beyond the boundary (Fig. 3.6), which leads to a decrease in the number of infectives. With LR dispersal, these edge effects are substantial (affecting up to 10% of the host population), while with SR dispersal they are quite small. Comparison of different system sizes and geometries shows that the stationary behaviour is considerably more affected by boundaries than by population size (Fig. 3.5). This suggests that epidemics in real populations, which are finite and have boundaries, might be smaller than predicted by standard models, especially when disease dispersal is long–ranged.

The results above refer to average behaviour; we also examined the variation of the mean epidemic size. The variance was found to increase with proximity to extinction, particularly with SR dispersal (Fig. 3.7).

The behaviour of the epidemic size could be understood in terms of the properties of the disease patterns (Fig. 3.31), which varied with the parameters in expected ways. We demonstrated that these properties are well characterised by the spatial autocorrelation function $C(r)$ (Cliff & Ord, 1981), which might therefore provide insight about the processes likely to have produced an observed disease pattern. The spatial correlation $C(r)$ is well fitted by stretched exponential functions (Fig. 3.13). The parameters of the fit suggest that $C(r)$ approaches a simple exponential for weakly–correlated patterns and a power–law for patterns with long–range correlations (i.e. closer to extinction) (Fig. 3.14). The LR dispersal case is exceptional in that $C(r)$, albeit much smaller than in the other cases, always exhibits a power–law form (Fig. 3.11). This feature might be expected

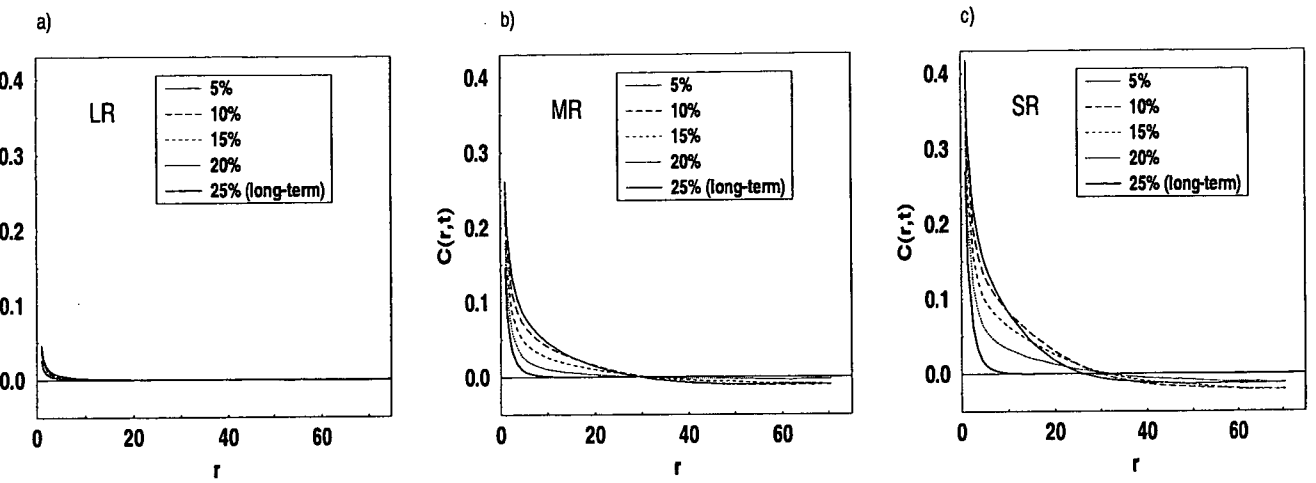


Figure 3.25: *Time evolution of the spatial correlation functions towards the stationary regime with* $\mu(\infty) = 25\%$ *for LR, MR and SR dispersal cases. Curves correspond to times when the epidemic first reached* 5%, 10%, 15% *and* 20% *and at equilibrium (top to bottom). Note that in the LR case correlations are very small and do not vary much with the disease level. a) LR; b) MR; c) SR.*

in the transient regime. However, since in the LR case epidemic size is well described by the MF model, which assumes a random pattern of disease, it might be thought that removals would destroy large–scale correlations in the stationary regime. This apparent contradiction is resolved by the fact that correlations are very small (Figs 3.9 and 3.10).

Transient behaviour. The essential point about transients is their stochastic variability, both within single epidemics and between epidemics under identical conditions (Kleczkowski *et al.*, 1996). This variability was illustrated in Figs 3.18 to 3.20 and Fig. 3.29. In practice, only single realisations of stochastic processes are usually observed (or multiple observations under variable conditions). However, due to epidemic variability, the study of a single disease progress curve might not be very useful for predicting the future of that epidemic or identical ones nor informative about its underlying processes. A more meaningful way of characterising transients and ascertaining their relation to underlying processes might be to use a large sample of identical realisations. We obtained this sample from simulated models, which would be very difficult in field experiments. We then used temporal and spatial statistics to identify the following relevant aspects of behaviour.

The time scale and shape of the mean disease progress curves $\mu(t)$ were found to vary considerably with the model parameters (Fig. 3.29). Excluding the case of LR dispersal (when disease patterns are fairly random), the curves are neither well

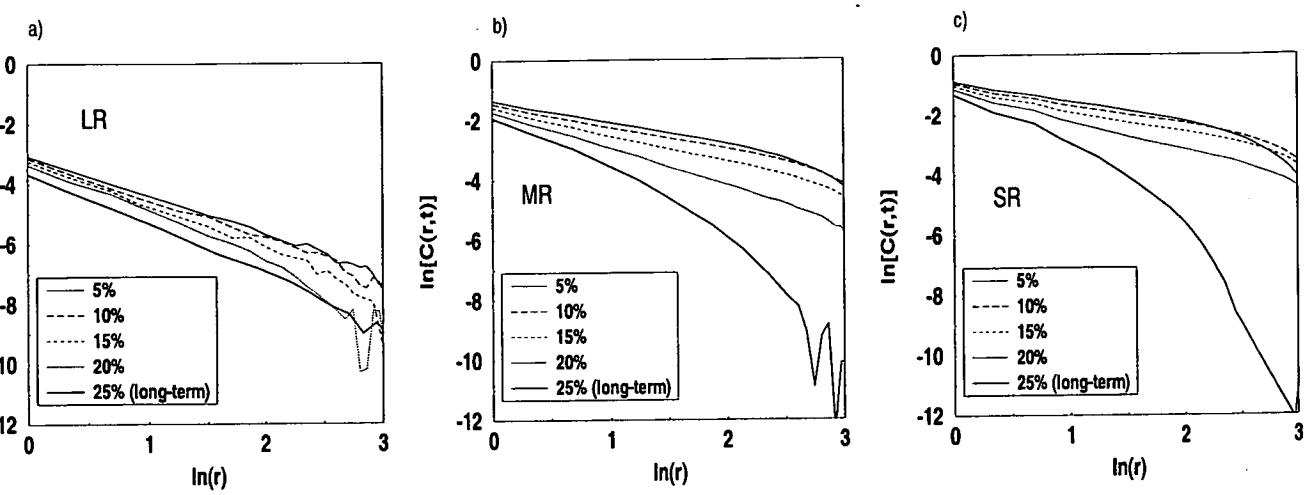


Figure 3.26: *Time evolution of the spatial correlation functions: ln–ln plots (over the range where $C(r,t) > 0$) corresponding to Fig. 3.25. a) LR; b) MR; c) SR.*

described by Mean Field predictions nor well fitted by standard logistic functions.

The corresponding standard deviation, $\sigma(t)$ (Fig. 3.20), contains additional information about the nature of the dispersal process. The shape of $\sigma(t)$, in particular its peak, was used for meaningfully defining the average transient duration t_{tr}. Whilst the definition has elements of arbitrariness (as do existing methods), it does capture the essential dependence on underlying processes (much better than standard descriptive methods) over a range of situations, including epidemics without recovery and different initial conditions. Definition of this time scale allowed rescaling and qualitative comparison of very different disease progress curves (Fig. 3.24).

We also analysed the time evolution of spatial autocorrelations. At any given time, their expectation over realisations, $C(r,t)$, splits into two regions separated by the same distance lag: a region of large, positive correlations (decaying in time to long–term values), indicative of complex, large–scale structure (as in Shaw, 1995), and a region of anticorrelations (vanishing in time) resulting from heterogeneity (Fig. 3.25). The occurrence of this type of anticorrelations in observed patterns could perhaps indicate a transient state. Transient correlation functions were found to have a power–law form in all dispersal cases (at least in the first region), even when long–term correlations did not (Fig. 3.26).

All the results and arguments presented in this chapter are qualitative and empirical since they rely on simulated samples. However, given the lack of observation data, and ambiguity in the information they provide, computer simulation is a valuable investigative tool. It enables us to gain insight into the underlying

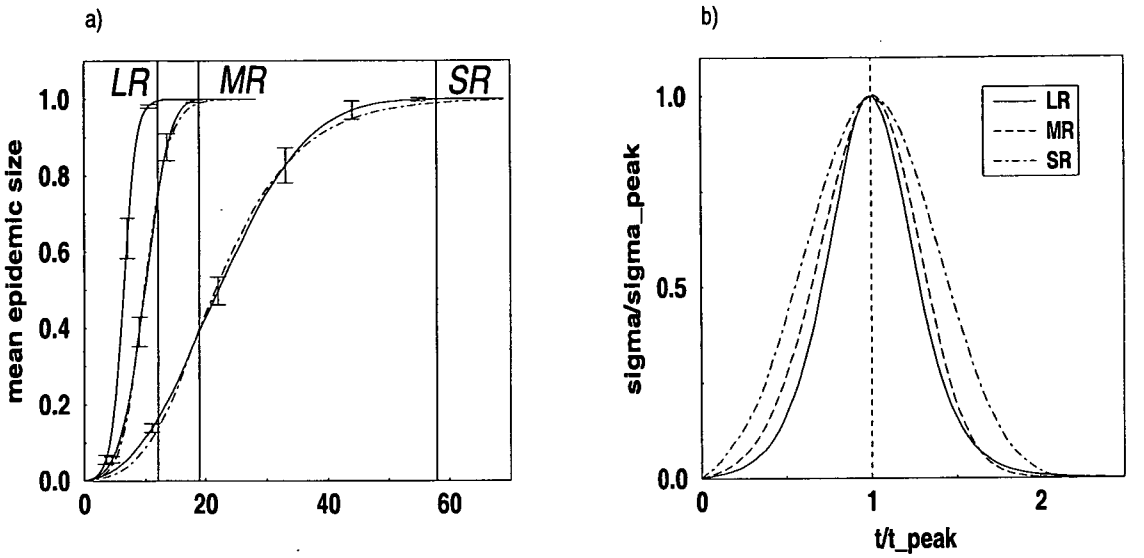


Figure 3.27: $R = 0$. *a) Mean of epidemic size over stochastic realisations, logistic fit and transient time scale (vertical line; as defined in section 3.6.4) for LR, MR and SR dispersal cases. Error bars show estimated standard deviations. MF prediction coincides with the LR curve. b) Standard deviation $\sigma(t)$ with axes rescaled by peak values $(\sigma/\sigma_p, t/t_p)$.*

mechanisms of epidemic development and provides a notable collection of results which can be used to develop and test analytical methods for general application. In chapter 6, various approaches designed to obtain deterministic predictions of expected behaviour will be tested through comparison with the results presented in this chapter.

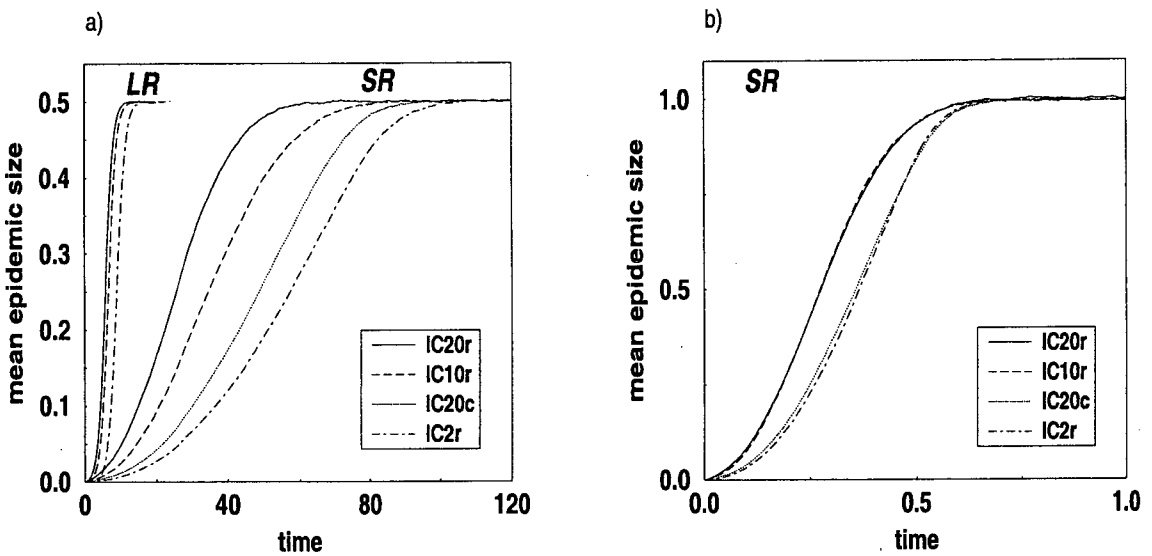


Figure 3.28: *Effect of initial conditions on average disease progress, for LR and SR dispersal, with* $\mu = 50\%$. *a)* $\mu(t)$; *b) SR case.* $\mu(t)$ *rescaled as in Fig. 3.24 with our estimate of* t_{tr} *(3.23).*

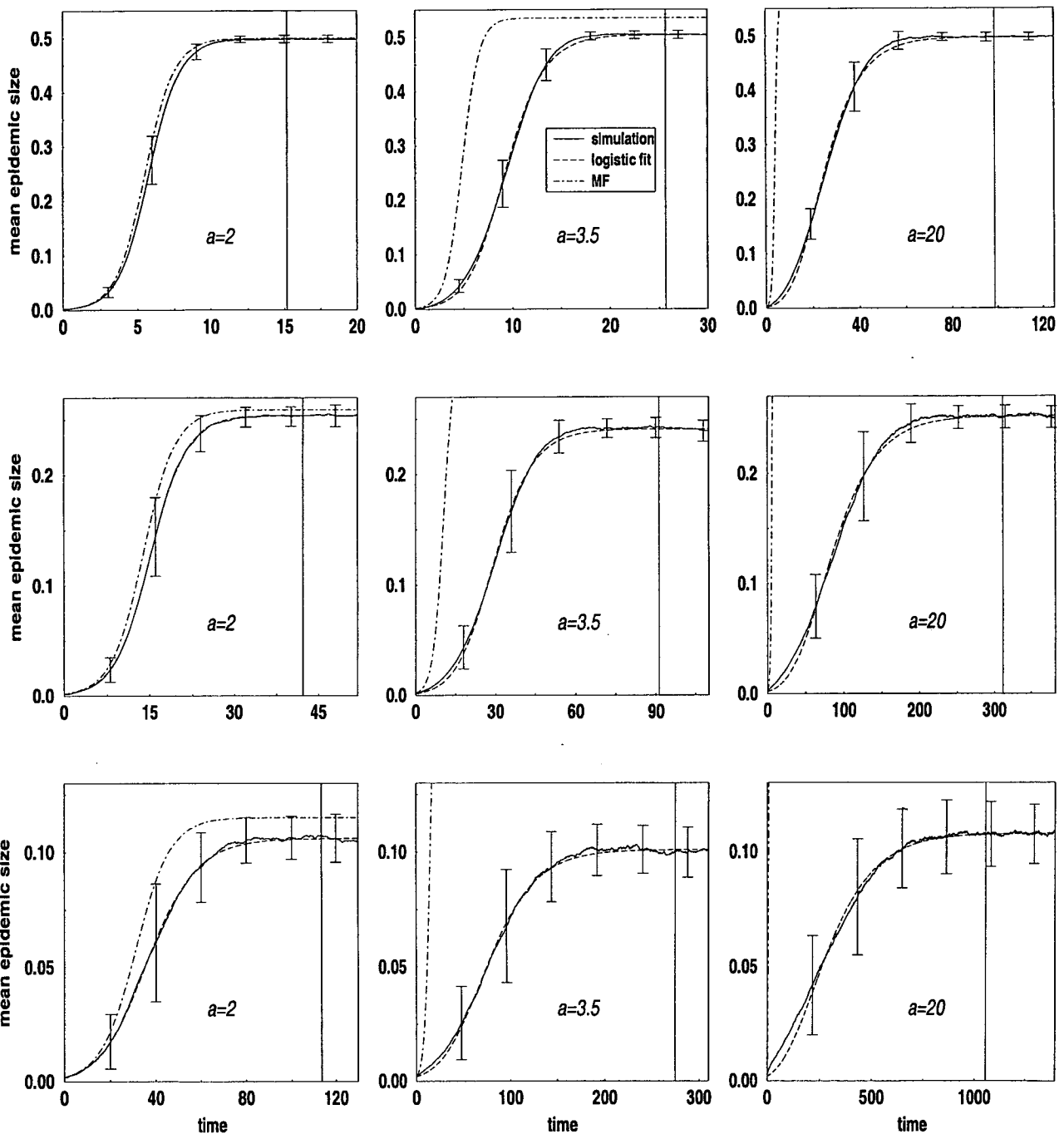


Figure 3.29: *Mean of epidemic size over stochastic realisations, logistic fit, Mean Field prediction and transient time scale (vertical line) as defined in section 3.6.2. Error bars show estimated standard deviations. From left to right: LR, MR, SR. From top to bottom:* $\mu = 50\%, 25\%, 10\%$.

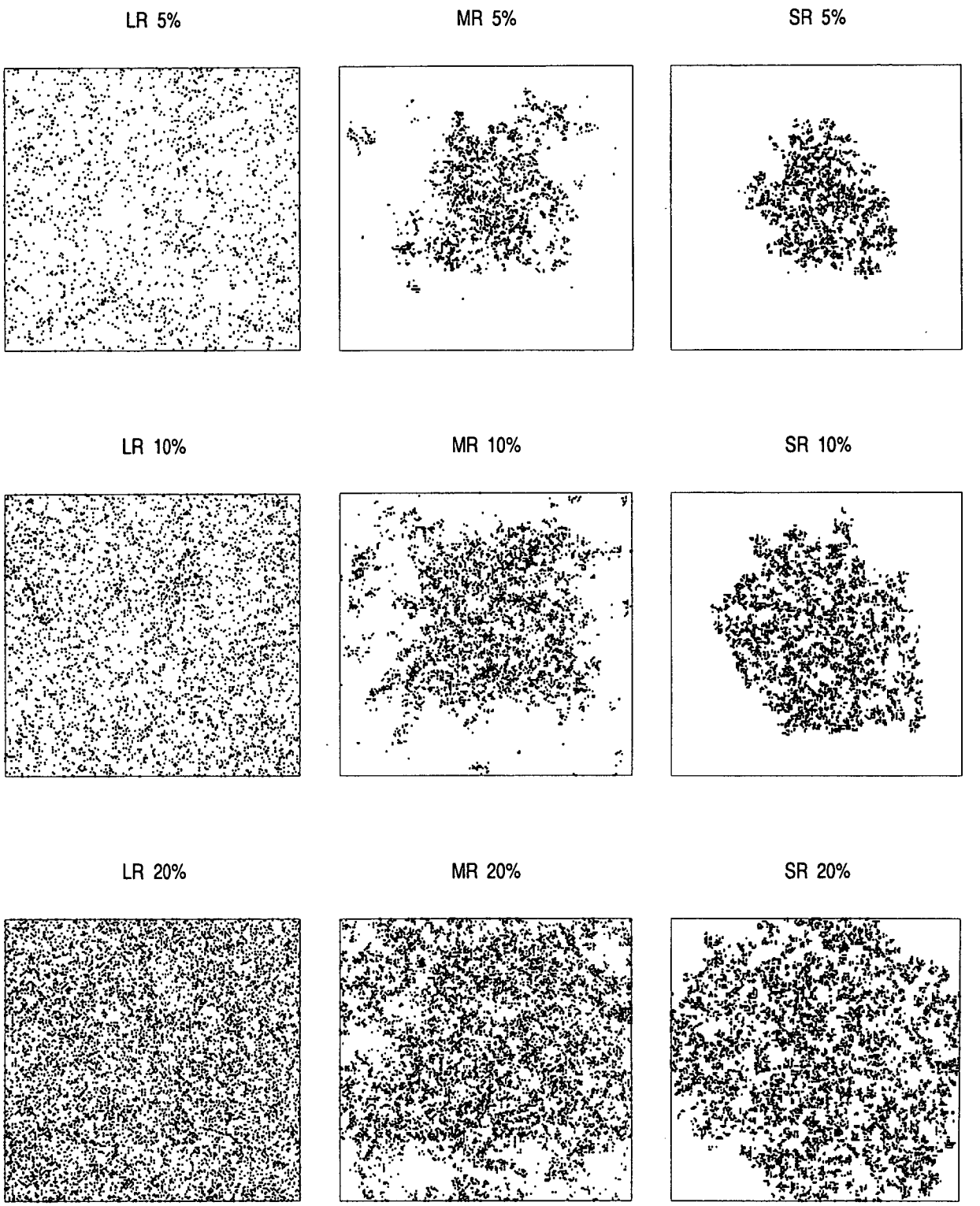


Figure 3.30: *Simulated disease patterns on a* 200×200 *square lattice with PBC. LR, MR and SR dispersal (left to right); transient states with* $5\%, 10\%$ *and* 20% *of infected hosts (top to bottom).*

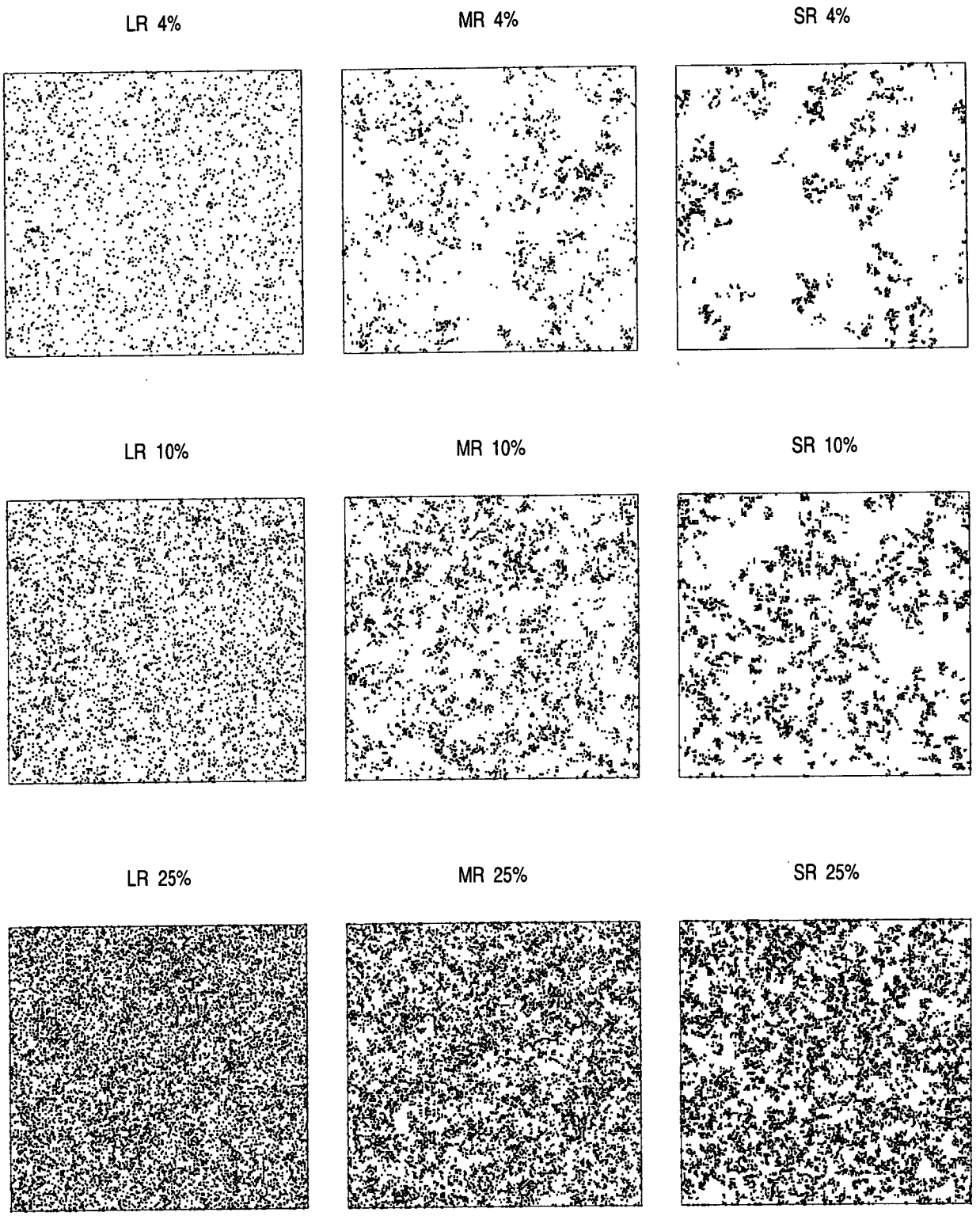


Figure 3.31: *Simulated disease patterns on a* 200×200 *square lattice with PBC. LR, MR and SR dispersal (left to right); stationary states with* $\mu = 4\%, 10\%$ *and* 25% *of infected hosts (top to bottom).*

Chapter 4

The nearest neighbour (NN) model

4.1 Introduction

In this chapter some of the existing analytical methods for the study of the behaviour of nearest neighbour models are reviewed, and their performance is assessed in the context of the plant epidemic model. In chapter 5 the method will be extended to an anisotropic version of this model.

The nearest neighbour (NN) model is a special case of the model under consideration, where the exponent a of the contact distribution is infinite and therefore interactions are restricted to neighbours at distance 1 (2 in 1-dimensional (1D) lattices, 4 in 2-dimensional (2D) square lattices). Models with nearest neighbour interactions have been extensively studied and used in many areas, such as mathematics, physics, chemistry and biology, to describe different kinds of interacting particle systems (Liggett, 1985, 1999; Konno, 1994). A well known example is the Contact Process (CP), introduced by Harris (1974) in probability theory, which, in our notation, corresponds to an epidemic process without background infection. NN models are also a special case of the stepping stone models used in population genetics (Kimura & Weiss, 1964).

Despite the simplicity of these models, exact solutions are very rare due to the intrinsic nonlinearity of interacting systems, and in particular no exact solutions are known for the CP in any spatial dimension. Simulation of stochastic spatio–temporal models such as the CP is computationally intensive because a large number of realisations might be needed to estimate any aspect of the system behaviour. It is therefore desirable to develop simpler analytical descriptions based on approximation techniques for predicting the model behaviour. Specifically, approaches based on cluster approximations have long been applied to studying models with NN interactions in statistical physics (e.g. Bethe, 1935; Kikuchi,

1951; ben-Avraham & Köhler, 1992) and more recently in population dynamics (e.g. Matsuda *et al.*, 1992; Sato *et al.*, 1994; Levin & Durrett, 1997; Filipe & Gibson, 1998, 2000).

We consider a succession of orders of cluster approximations: Mean Field (MF, first order), Pairwise Approximation (PA, second order; see: Dickman, 1986; Sato *et al.*, 1994; Levin & Durrett, 1997; Filipe & Gibson, 1998) and Hybrid Pairwise Approximation (HPA; Filipe, 1999 a), which is a modified version of the second order PA. Higher order approximations are commonly used for 1D systems (ben-Avraham & Köhler, 1992; Bassler & Browne, 1996, 1998) but they have been applied to 2D systems to a much lesser extent because in 2 dimensions equations become much more complex with increasing order of approximation (Kramers & Wannier, 1941; Hiebeler, 1997; Filipe & Gibson, 2000). We apply third and fourth–order approximations (3A, 4A) to the 1D model and the third–order approximation (Squarewise Approximation, SA) to the 2D model (Filipe & Gibson, 2000). Predictions for the model behaviour are then compared to the other lower–order approximations and to simulation.

4.2 NN dynamic equations and cluster approximations

For interactions restricted to NNs on a generic lattice with z neighbours at each site, equations (2.21) and (2.22) become:

$$\frac{dP[1]}{dt} = J_1 - (J_1 - zJ + R)P[1] - zJP[11] \tag{4.1}$$

$$\frac{1}{2}\frac{dP[11]}{dt} = (J_1 + J)\,P[1] - (J_1 + J + R)P[11] + J\sum_{\mathbf{v}'}\langle S_{\mathbf{x}}(1 - S_{\mathbf{y}})S_{\mathbf{v}'}\rangle, \tag{4.2}$$

where $J = J_2/z$ is the rate at which pathogens are deposited on each NN of an infected site, $\mathbf{y}$ is a NN of $\mathbf{x}$, and the $\mathbf{v}'$–sum runs over the $z-1$ NNs of $\mathbf{y}$ distinct from $\mathbf{x}$. For clarity, we are now using a slightly different notation from the one adopted in chapter 2: here, $P[1]$ indicates the probability that a site is infected at time t ($P[1] = P_I(t)$); $P[ab]$ indicates the probability that two neighbouring sites are in the states a and b at time t (e.g. $P[11] = P_{II}(1,t)$ is the probability that two neighbours are infected, assuming a lattice spacing of 1; $P[01]$ is the probability that neighbouring sites are in different states, susceptible and infected). The term $\sum_{\mathbf{v}'}\langle S_{\mathbf{x}}(1-S_{\mathbf{y}})S_{\mathbf{v}'}\rangle$ refers to triplets of NN sites. For example, on a square lattice

$(z = 4)$

$$\sum_{\mathbf{v}'} \langle S_{\mathbf{x}}(1 - S_{\mathbf{y}}) S_{\mathbf{v}'} \rangle = P[101] + 2\, P\begin{bmatrix} 1 & 0 \\ & 1 \end{bmatrix}, \tag{4.3}$$

where $P[abc]$, $P\left[\begin{smallmatrix} a & b \\ & c \end{smallmatrix}\right]$ refer to triplets of adjacent sites, either collinear or at right angle, in states a, b, c $(a, b, c \in \{0, 1\})$, at time t. In order to simplify the notation, the dependence on time is not indicated explicitly. We recall that equations (2.21) and (2.22), and consequently equations (4.1) and (4.2), describe the dynamics of a system which is spatially–homogeneous and isotropic at all times.

Like equations (2.21) and (2.22), equations (4.1) and (4.2) are the first two of an unlimited hierarchy of equations for the densities of consecutively larger clusters in which any subsystem is not closed: the equation for $P[1]$ depends on $P[11]$, the equation for $P[11]$ involves triplets of adjacent sites, and so on.

Cluster approximations are used to truncate the system and obtain approximate solutions. The idea is that spatial correlations inside a basic cluster are fully accounted for, while simplifying assumptions are made regarding correlations between individuals inside and outside a cluster. In this way, densities of clusters larger than the basic cluster can be expressed in terms of densities of smaller clusters, thereby closing the system of equations (Filipe & Gibson, 1998).

The basic cluster is defined by the maximum distance within which full correlations are accounted for. On a 1D linear lattice, the possible distances between sites are the non–negative integers, so cluster definition is trivial. On a square lattice the distances are: $0, 1, \sqrt{2}, 2$, etc. and the corresponding basic clusters are: a site, a pair of adjacent sites, a square of side 1, a cross comprising a central site and its four NNs, etc. For larger clusters, the number of equations needed and the number of terms which they comprise become too large for the analysis to be tractable.

Low–order cluster approximations are relatively straightforward to implement and yield good descriptions of behaviour, at least in parameter regions where spatial correlations are not too large. It is expected that the accuracy of the approximation increases as the size of the basic cluster increases. This convergence has been demonstrated for cluster approximations up to order 10 in 1D systems (ben-Avraham & Köhler, 1992) but only up to the third order in 2D systems (Filipe & Gibson, 2000), for which the equations become cumbersome beyond the second oder.

Figs 4.1 and 4.2 compare simulation and predictions of the different cluster approximations for the stationary epidemic size.

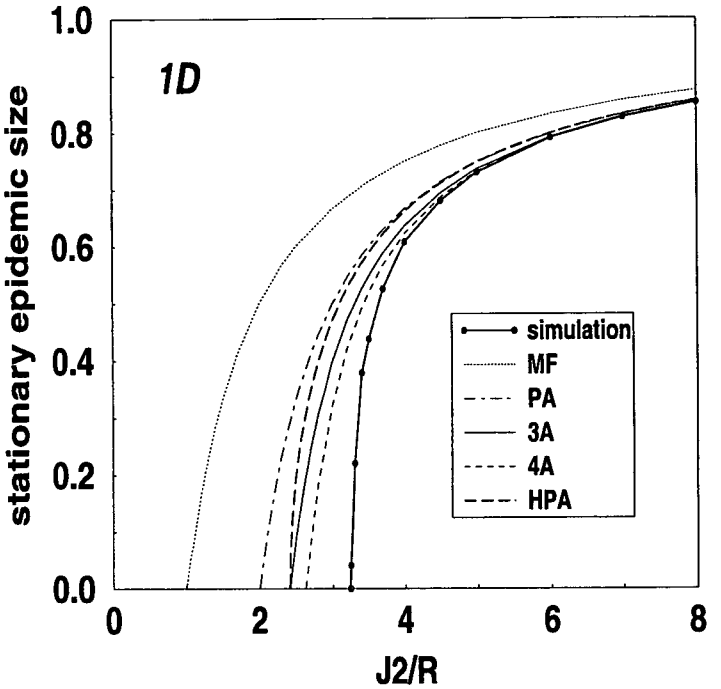


Figure 4.1: *Stationary epidemic size obtained from simulation and predicted by cluster approximations (MF, PA, 3A, 4A, HPA) on a 1D lattice.*

4.2.1 Mean Field (MF)

Mean Field (MF) is the simplest and the lowest–order cluster approximation. MF assumes that the states of any two individuals in the system are independent of each other, i.e. that there are no spatial correlations. This means that

$$P[11] = P[1]^2, \tag{4.4}$$

and the basic cluster consists of a single site. The first equation of system (4.1)–(4.2) then becomes

$$\frac{dP[1]}{dt} = J_1 - (J_1 + R)\ P[1] + zJP[1]\ (1 - P[1]). \tag{4.5}$$

This is a logistic equation whose explicit solution is given by (e.g. Filipe & Gibson, 1998)

$$P[1](t) = \frac{\gamma \tanh\ (\gamma\ t/2 + C) + A}{2\ zJ}, \tag{4.6}$$

where

$$A = zJ - R - J_1, \tag{4.7}$$

$$\gamma = \sqrt{A^2 + 4z\ J_1\ J}, \tag{4.8}$$

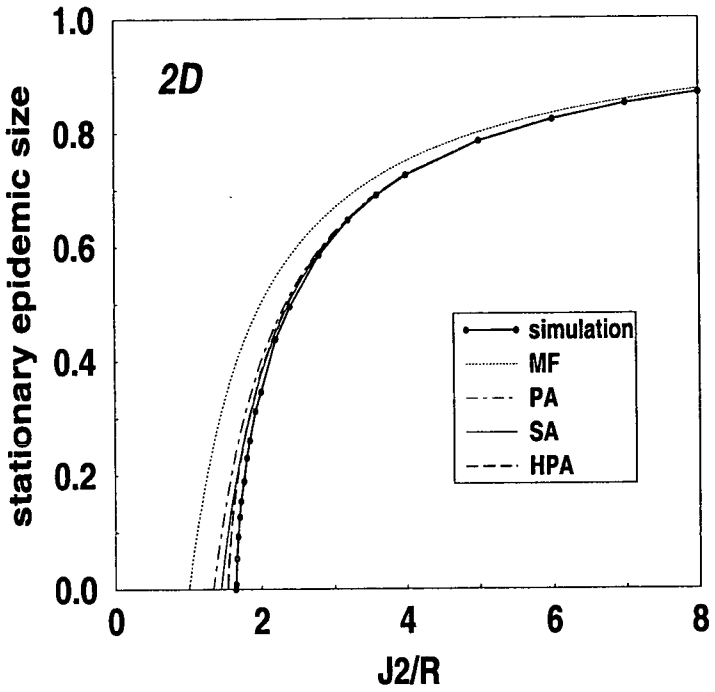


Figure 4.2: *Stationary epidemic size obtained from simulation and predicted by cluster approximations (MF, PA, SA, HPA) on a 2D lattice.*

$$C = \tanh^{-1}\left(\frac{2\ zJ\ P_0 - A}{\gamma}\right), \tag{4.9}$$

and P_0 is the initial density of infectives. The long–term, equilibrium level of infection is

$$P[1]_\infty = \frac{\gamma + A}{2\ zJ}. \tag{4.10}$$

MF predicts that in the absence of background infection ($J_1 = 0$) the system undergoes a phase transition from a parameter region where the epidemic persists to another where it dies out. When $J_1 = 0$, (4.6) reads (Filipe & Gibson, 1998)

$$P[1](t) = \frac{P_0}{\exp[(R - zJ)\ t] + \mathrm{sgn}(R - zJ)[\exp[(R - zJ)\ t] - 1]P_0/P[1]_\infty}, \tag{4.11}$$

and the long–term level of infection is

$$P[1](\infty) = \begin{cases} 0 & R \geq zJ \\ 1 - R/(zJ) & R < zJ \end{cases}. \tag{4.12}$$

Hence, the MF prediction for the epidemic persistence threshold is $(J/R)_c = 1/z$ (see Figs 4.1, 4.2).

Since MF ignores spatial correlations, this approximation can only provide good predictions when the distribution of infectives and susceptibles is nearly

random; in other words, when there is a high degree of mixing between susceptibles and infectives. When correlations are large or long–ranged we expect the system to deviate significantly from the MF ideal; this has been observed in the simulation studies of the model (see chapter 3; Filipe & Gibson, 1998; Filipe *et al.*, 2000; Filipe & Maule, 2000 a). Substituting the autocorrelation function at spatial lag 1 (see (2.19)),

$$C(1) = \frac{P[11] - P[1]^2}{P[1](1 - P[1])}, \tag{4.13}$$

into equation (4.1) gives

$$\frac{dP[1]}{dt} = J_1 - (J_1 + R)P[1] + zJP[1]\ (1 - P[1])\ (1 - C(1)). \tag{4.14}$$

Similarly, for a model with general interactions, it was shown in chapter 2 that the deviation from MF is given by a convolution of the spatial correlation and the contact distribution (2.30). Therefore, when correlations are not negligible, MF cannot provide accurate predictions and higher order approximations need to be considered.

4.2.2 Pairwise Approximation (PA)

The next order of approximation is the Pairwise Approximation (PA), which accounts for full correlations between NNs, i.e. the basic cluster is a pair of NN sites (Filipe & Gibson, 1998). Specifically, PA assumes that the states of two sites x and z with a common NN, y, are independent conditional on the state of y,

$$P[xyz] = P[x|yz]\ P[yz] \approx P[x|y]\ P[yz] = \frac{P[xy]}{P[y]}\ P[yz]. \tag{4.15}$$

This implies that

$$P[101] = P\begin{bmatrix} 1 & 0 \\ & 1 \end{bmatrix} = \frac{P[10]^2}{P[0]}. \tag{4.16}$$

Since

$$P[0] = 1 - P[1] \quad \text{and} \quad P[10] = P[1] - P[11], \tag{4.17}$$

equations (4.1) and (4.2) can be rewritten as

$$\frac{dP[1]}{dt} = J_1 - (J_1 - zJ + R)\ P[1] - zJ\ P[11] \tag{4.18}$$

$$\frac{1}{2}\frac{dP[11]}{dt} = (J_1 + J)\ P[1] - (J_1 + J + R)\ P[11] + (z-1)J\ \frac{(P[1] - P[11]^2)}{1 - P[1]}. \quad (4.19)$$

It is a straightforward task to solve this system of nonlinear equations numerically for given initial conditions. Equilibrium levels of disease are obtained by setting time derivatives to zero. From (4.18) we have

$$P[11]_\infty = \frac{J_1 - (J_1 - zJ + R)}{zJ} P[1]_\infty, \quad (4.20)$$

which, substituted in (4.19), gives an equation in $P[1]_\infty$ with a single solution satisfying $0 \leq P[1] \leq 1$ (Filipe & Gibson, 1998).

In the absence of background infection ($J_1 = 0$), the solution is

$$P[1]_\infty = \frac{z\ (z - 1 - R/J)}{z\ (z-1) - R/J}, \quad \text{for } J/R > (J/R)_c, \quad (4.21)$$

and $P[1]_\infty = 0$ otherwise, where $(J/R)_c = 1/(z-1)$ (Filipe & Gibson, 1998).

PA significantly improves MF predictions for both stationary and transient behaviour (Figs 4.1, 4.2). However, there is still a considerable mismatch between PA predictions and simulation, especially near the persistence threshold. This is explained by the fact that in the proximity of the phase transition long–range spatial correlations develop in the system which cannot be captured by PA or other cluster approximations. In fact, it has been shown (Filipe, 2000, unpublished) that PA's assumption of conditional independence implies that correlations decay exponentially with distance, i.e.

$$C(r) = C(1)^r, \quad (4.22)$$

where $C(r)$ is the autocorrelation function at spatial lag r and r is the norm 1 distance (also known as Manhattan metric) defined as follows. Given two points with Cartesian coordinates (x_1, y_1) and (x_2, y_2), their norm 1 distance r is given by

$$r = |x_1 - x_2| + |y_1 - y_2|. \quad (4.23)$$

We reproduce the detailed proof of (4.22) here. First we consider a 1D linear lattice. In this case, the norm 1 distance and the Euclidean distance coincide. We start by proving the result for $r = 2$, i.e.

$$C(2) = C(1)^2. \quad (4.24)$$

The probability that two sites at distance $r = 2$ are infected, $P_{II}(2)$, can be written in the form

$$P_{II}(2) = P[1 \bullet 1] = P[111] + P[101], \tag{4.25}$$

where $\bullet$ denotes a site in an unspecified state. Applying (4.15) and using (4.17) yields

$$P_{II}(2) = \frac{P[11]^2\ (1 - P[1]) + P[1]\ (P[1] - P[11])^2}{P[1]\ (1 - P[1])}. \tag{4.26}$$

Adding and subtracting $P[1]^4$ from the numerator gives

$$P_{II}(2) = \frac{(P[11] - P[1]^2)^2}{P[1]\ (1 - P[1])} + P[1]^2. \tag{4.27}$$

The autocorrelation at spatial lag 2 is given by

$$C(2) = \frac{P_{II}(2) - P[1]^2}{P[1]\ (1 - P[1])}. \tag{4.28}$$

Substituting (4.27) in (4.28) finally gives

$$C(2) = \left(\frac{P[11] - P[1]^2}{P[1]\ (1 - P[1])} \right)^2 = C(1)^2. \tag{4.29}$$

We now use induction to prove (4.22) for any spatial lag. We assume that $C(r) = C(1)^r$ and wish to show that this implies

$$C(r + 1) = C(1)^{r+1}. \tag{4.30}$$

Consider the probability that two sites at distance $r + 1$ are both infected, regardless of the states of the r sites between them,

$$P_{II}(r + 1) = P[\underbrace{1 \cdots 1}_{r+2}]. \tag{4.31}$$

This can be rewritten as

$$P_{II}(r + 1) = P[\underbrace{1 \cdots 1\ 1}_{r+2}] + P[\underbrace{1 \cdots 0\ 1}_{r+2}]. \tag{4.32}$$

Applying (4.15) to the triplets of non–adjacent sites $P[\underbrace{1 \cdots 1\ 1}_{r+2}]$ and $P[\underbrace{1 \cdots 0\ 1}_{r+2}]$ yields

$$P[\underbrace{1 \cdots 1\ 1}_{r+2}] = \frac{P[\overbrace{1 \cdots 1}^{r+1}]\ P[11]}{P[1]}, \tag{4.33}$$

$$P[\underbrace{1\cdots0\ 1}_{r+2}] = \frac{P[\overbrace{1\cdots0}^{r+1}]\ P[10]}{P[0]}, \tag{4.34}$$

i.e. we assume that the states of the site in the first position and the site in position $r+2$ (at distance $r+1$ from each other) are independent conditional on the state of the site in position $r+1$. We denote $P[\underbrace{1\cdots0}_{r+1}]$ (the probability that in a pair of sites at a distance r the first site is infected and the second is susceptible) by $P_{IS}(r)$ and use

$$P_{IS}(r) = P[1] - P_{II}(r). \tag{4.35}$$

Substituting (4.33), (4.34) and (4.35) in (4.32), and using (4.17), gives

$$P_{II}(r+1) = \frac{P_{II}(r)\ P[11]}{P[1]} + \frac{(P[1] - P_{II}(r))\ (P[1] - P[11])}{1 - P[1]}. \tag{4.36}$$

Using this in the definition of $C(r+1)$,

$$C(r+1) = \frac{P_{II}(r+1) - P[1]^2}{P[1]\ (1 - P[1])}, \tag{4.37}$$

gives

$$C(r+1) = \frac{(P_{II}(r) - P[1]^2)}{(P[1]\ (1 - P[1]))}\ \frac{(P[11] - P[1]^2)}{(P[1]\ (1 - P[1]))} \tag{4.38}$$

which is equal to

$$C(r+1) = C(r)\ C(1) = C(1)^{r+1} \tag{4.39}$$

and completes the proof of (4.22).

We have shown that on a 1D lattice PA implies exponential decay of spatial correlations. The same result holds for the autocorrelations at any *integer* spatial lag on a 2D square lattice. The proof can be extended to non–integer spatial lags in 2D systems following an argument analogous to the one given for the 1D case (Filipe, 2000, unpublished).

4.2.3 Higher order approximations

Approximations above the pairwise order are formulated differently in 1D than 2D lattices. This is because the basic cluster is defined by the maximum distance at which correlations are fully accounted for, and in 2D there are non–integer distances between sites. In 1D systems, the third–order approximation involves

clusters of three sites (at maximum distance 2), the fourth–order approximation involves clusters of four sites (at maximum distance 3), and so on. On 2D lattices, the relevant distance to the third order of approximation is $\sqrt{2}$ (not 2) and the basic cluster is a square of side 1.

Higher order approximations have been widely used in statistical physics, for example in the context of surface reaction systems for 1D models (e.g. ben-Avraham & Köhler, 1992; Bassler & Browne, 1996, 1998). Here, we apply third and fourth–order approximations (3A and 4A) to our NN epidemic model realised on a 1D lattice. Recently, Filipe & Gibson (2000) proposed an approximation for 2D systems which accounts for correlations up to distance $\sqrt{2}$ (Squarewise Approximation, SA). We test the performance of 3A, 4A and SA in the context of our NN epidemic model realised on 1D and 2D lattices by comparing their predictions to other approximations and simulation results. We shall consider the 1D and 2D cases separately in the following section.

4.2.3.1 1D lattice. Third and fourth–order approximations (3A, 4A)

On 1D lattices, the number of neighbours at each site is $z = 2$. Equation (4.1) is unchanged but equation (4.2) reads

$$\frac{1}{2}\frac{dP[11]}{dt} = (J_1 + J)\,P[1] - (J_1 + J + R)P[11] + JP[101]. \qquad (4.40)$$

PA closes this system of equations by assuming the relation (4.16) between $P[101]$ and $P[11]$ and $P[1]$. In order to go one step further, a dynamic equation for the density $P[101]$ must be derived.

This is done by considering all possible infection and recovery events which might create or destroy a triplet of adjacent sites [101]. A triplet [101] can be created through:

- primary infection of the first site in the triplet [001] or of the third site in the triplet [100],
- recovery of the central site of the triplet [111],
- transmission of disease from the first to the second site or from the fourth to the third site in the quadruplet [1001];

and can be destroyed through:

- recovery of the first or third site of the triplet [101],
- primary infection of the central site in the triplet [101],

- transmission of disease from the first or third site to the central site of the triplet [101].

This leads to the following equation

$$\frac{dP[101]}{dt} = 2J_1P[100] + RP[111] + 2JP[1001] - (J_1 + 2R + 2J)P[101]. \quad (4.41)$$

This equation involves the triplet of sites [111] and hence another equation for the density $P[111]$ must be derived. This is done along the same lines as above, and gives

$$\frac{dP[111]}{dt} = (J_1 + 2J)P[101] + 2(J_1 + J)P[110] + 2JP[1011] - 3RP[111]. \quad (4.42)$$

These two equations involve densities of clusters of four sites in different states. In order to close the system of equations at the third order, we assume that sites at distance larger than 2 are independent, conditional on the states of the intermediate sites, i.e.

$$P[abcd] = P[a|bcd]\ P[bcd] \approx P[a|bc]\ P[bcd] = \frac{P[abc]\ P[bcd]}{P[bc]}, \quad (4.43)$$

where $a, b, c, d \in \{0, 1\}$. To close the system at a higher order, equations for the densities of 4–site clusters must also be derived, which will in turn involve clusters with 5 sites. Then, closure at fourth order assumes conditional independence between sites at distance larger than 3, and so on.

The third order of approximation (3A) leads to the system of equations given by (4.1), (4.40)–(4.42), where

$$P[1001] \approx \frac{P[100]^2}{P[00]} \quad (4.44)$$

and

$$P[1011] \approx \frac{P[101]\ P[110]}{P[10]}. \quad (4.45)$$

In addition we use the exact relations: (4.17), $P[00] = P[0] - P[10]$, $P[100] = P[10] - P[101]$ and $P[110] = P[11] - P[111]$.

The fourth order of approximation (4A) leads to a system of seven equations: the previous four equations plus three equations for the densities $P[1001]$, $P[1011]$ and $P[1111]$, which we do not report here since they are cumbersome and can be easily derived in much the same way as equation (4.41).

Fig. 4.1 shows how cluster approximations of different order (from MF to 4A) perform in predicting the simulated stationary level of infection for a 1D epidemic model. As expected, the higher the order of approximation, the closer the prediction is to the simulation curve, though the improvement of each order of approximation relative to the previous order decreases as the order increases. PA performs much better than MF, while 4A gives only slightly better predictions than 3A. However, all approximations fail to predict the infinite slope of the curve in the proximity of the persistence threshold.

4.2.3.2 2D lattice. Squarewise approximation (SA)

On the plane, the maximum distance at which spatial correlations are accounted for in the second order of approximation is $\sqrt{2}$, hence the basic cluster is the square lattice cell (Filipe & Gibson, 2000).

In this case, there are three additional independent density variables and thus three extra equations are needed relative to PA. At second order (PA), it does not matter whether the second equation, (4.2), is on $P[11]$, $P[10]$ or $P[00]$ since application of the PA closure (4.15) to any equation leads to an equivalent system of equations. At third order (or higher), there are different ways of closing the system of equations, depending on the choices of independent variables and of higher–order densities which are decoupled using the assumption of conditional independence. Like Filipe & Gibson (2000), we have considered five equations, on either densities of clusters which only contained infected sites or densities of clusters which only contained susceptible sites, and some hybrid cases. We have verified that, although the corresponding systems of equations are not equivalent after decoupling, the numerical solutions they produce are very similar.

Here, in addition to equation (4.1) and (4.2), we formulate the following equations:

$$\begin{aligned}\frac{dP\left[\begin{smallmatrix}1\\0\ 1\end{smallmatrix}\right]}{dt} = {} & R\,P\begin{bmatrix}1\\1\ 1\end{bmatrix} - (J_1+2J+2R)\,P\begin{bmatrix}1\\0\ 1\end{bmatrix} + 2J_1\,P\begin{bmatrix}0\\0\ 1\end{bmatrix} \\ & + 2J\left(P\begin{bmatrix}1\\0\\0\ 1\end{bmatrix} + P\begin{bmatrix}1\ 0\\0\ 1\end{bmatrix} + P\begin{bmatrix}1\ 1\\0\ 0\end{bmatrix} - P\begin{bmatrix}1\\1\ 0\ 1\end{bmatrix}\right), \end{aligned} \tag{4.46}$$

$$\begin{aligned}\frac{dP\left[\begin{smallmatrix}1\\1\ 1\end{smallmatrix}\right]}{dt} = {} & (J_1+2J)\,P\begin{bmatrix}1\\0\ 1\end{bmatrix} + 2(J_1+J)\,P\begin{bmatrix}0\\1\ 1\end{bmatrix} - 3R\,P\begin{bmatrix}1\\1\ 1\end{bmatrix} \\ & + 2J\left(P\begin{bmatrix}1\\1\ 0\\1\end{bmatrix} + P\begin{bmatrix}1\\1\ 0\ 1\end{bmatrix} + P\begin{bmatrix}1\\1\ 0\ 1\end{bmatrix} + P\begin{bmatrix}1\ 0\\1\ 1\end{bmatrix}\right) \end{aligned} \tag{4.47}$$

and

$$\frac{1}{4}\frac{dP\left[\begin{smallmatrix}1&1\\1&1\end{smallmatrix}\right]}{dt} = -R\,P\begin{bmatrix}1&1\\1&1\end{bmatrix} + (J_1+2J)\,P\begin{bmatrix}1&0\\1&1\end{bmatrix} + 2J\,P\begin{bmatrix}1&\\0&1\\1&1\end{bmatrix}. \tag{4.48}$$

The system of equations is closed using the assumption of conditional independence between sites at distance larger than $\sqrt{2}$ (Filipe & Gibson, 2000), which translates into:

$$P[101] \approx \frac{P[10]^2}{P[0]} \tag{4.49}$$

$$P\begin{bmatrix}1&\\0&\\0&1\end{bmatrix} \approx \frac{P\left[\begin{smallmatrix}0\\0&1\end{smallmatrix}\right]\,P[10]}{P[0]} \tag{4.50}$$

$$P\begin{bmatrix}1&0\\&0&1\end{bmatrix} \approx \frac{P\left[\begin{smallmatrix}0\\0&1\end{smallmatrix}\right]^2}{P[00]} \tag{4.51}$$

$$P\begin{bmatrix}&1&\\1&0&1\end{bmatrix} \approx \frac{P\left[\begin{smallmatrix}1\\0&1\end{smallmatrix}\right]^2}{P[10]} \tag{4.52}$$

$$P\begin{bmatrix}1&\\1&0\\&1\end{bmatrix} \approx \frac{P\left[\begin{smallmatrix}1\\1&0\end{smallmatrix}\right]\,P\left[\begin{smallmatrix}1\\0&1\end{smallmatrix}\right]}{P[10]} \tag{4.53}$$

$$P\begin{bmatrix}1&\\1&0&1\end{bmatrix} \approx \frac{P\left[\begin{smallmatrix}1\\1&0\end{smallmatrix}\right]\,P[10]}{P[0]} \tag{4.54}$$

$$P\begin{bmatrix}1&\\0&1\\1&1\end{bmatrix} \approx \frac{P\left[\begin{smallmatrix}1\\0&1\end{smallmatrix}\right]\,P\left[\begin{smallmatrix}0&1\\1&1\end{smallmatrix}\right]}{P[10]}. \tag{4.55}$$

In addition, we use the exact relations: (4.17), $P[00] = P[0] - P[10]$, $P\left[\begin{smallmatrix}0\\0&1\end{smallmatrix}\right] = P[10] - P\left[\begin{smallmatrix}1\\0&1\end{smallmatrix}\right]$, $P\left[\begin{smallmatrix}1\\1&0\end{smallmatrix}\right] = P[11] - P\left[\begin{smallmatrix}1\\1&1\end{smallmatrix}\right]$, $P\left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right] = P\left[\begin{smallmatrix}1\\1&1\end{smallmatrix}\right] - P\left[\begin{smallmatrix}1&1\\1&1\end{smallmatrix}\right]$, $P\left[\begin{smallmatrix}1&1\\0&0\end{smallmatrix}\right] = P\left[\begin{smallmatrix}1\\1&0\end{smallmatrix}\right] - P\left[\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right]$.

We solved these equations numerically and found, as expected, that SA performs better than PA in all parameter regions, even though the improvement is modest, especially in the critical region. Fig. 4.2 compares the phase diagrams from the approximation and simulation. As already observed in 1D systems, even though higher order approximations seem to converge to the simulated phase diagram, the slope of the curves is wrong: cluster approximations are unable to predict the infinite slope of the curve at the critical threshold. This is a consequence of their inability to capture large scale fluctuations which develop in the proximity of the critical point (Filipe & Gibson, 1998; Filipe *et al.*, 2000).

4.2.4 Hybrid Pairwise Approximation (HPA)

Recently, Filipe (1999 a) has devised an elaboration of PA which tries to overcome this deficiency and provides very accurate predictions over the whole parameter space, including the critical region. The new approach is examines in this section, which is a summary of the results in Filipe (1999 a, b) .

The underlying idea for developing this new approach was the observation that cluster approximations tend to overestimate the rate of progress and the asymptotic size of epidemics. Since the approximations are based on neglecting some correlations, we intuitively expect the overestimation of the disease level to be a consequence of underestimating the amount of local clustering between infecteds. Equation (4.14) confirms this assertion: smaller values of $C(1)$ lead to larger values of $P[1]$. MF, for example, underestimates $P[11]$ by replacing it with $P[1]^2$; since $P[11]$ appears in equation (4.1) with a negative sign, the MF prediction for $P[1]$ is overestimated. Similarly, PA underestimates the triplet terms $P[101], P\left[\begin{smallmatrix}1\\0\,1\end{smallmatrix}\right]$ in equation (4.2), leading to an underestimation of $P[11]$, which in turn causes an overestimation of $P[1]$ in equation (4.1).

The Hybrid Pairwise Approximation (HPA) is formulated for a more general epidemic model than the one considered here (Filipe, 1999 a, b). It requires consideration of a model in which the recovery process is as follows: infectives are still replaced by susceptibles at rate R but, in addition, all infected NNs of a recovered individual are replaced by susceptibles with probability α, where $0 \leq \alpha \leq 1$. Our model simply corresponds to the case $\alpha = 0$. For general α, equations (4.1) and (4.2) become

$$\frac{dP[1]}{dt} = J_1 - (J_1 - zJ + R)P[1] - z(J + R\alpha)\ P[11] \qquad (4.56)$$

$$\frac{1}{2}\frac{dP[11]}{dt} = (J_1 + J)\ P[1] - (J_1 + J + R)\ P[11] + J\left(P[101] + (z-2)\ P\begin{bmatrix}1\\0\ 1\end{bmatrix}\right) - R\alpha\left(P[111] + (z-2)\ P\begin{bmatrix}1\\1\ 1\end{bmatrix}\right). \quad (4.57)$$

PA approximates the triplet terms in (4.57) as

$$J\ (z-1)\ (P[1] - P[11])^2/(1 - P[1]) - R\alpha\ (z-1)\ P[11]^2/P[1]. \quad (4.58)$$

Since the two triplet terms have different signs and their magnitudes are underestimated, it is not obvious what the overall effect of PA on $dP[11]/dt$ (given by (4.57)), and thus on $dP[1]/dt$, is when $\alpha > 0$. The fact that PA overestimates $P[1]$ for $\alpha = 0$ as well as for $\alpha = 1$ (Filipe & Gibson, 1998; Filipe, 1999 a) suggests that the sum of triplet terms is underestimated for $0 \leq \alpha \leq 1$.

The presence of two triplet terms in (4.57) with opposite signs has suggested the construction of a variant of PA which underestimates $P[1]$. This was done by approximating the negative triplet term as follows

$$-R\alpha\left(P[111] + (z-2)\ P\begin{bmatrix}1\\1\ 1\end{bmatrix}\right) \quad (4.59)$$

$$= -R\alpha\left(P[1 \bullet 1] + (z-2)\ P\begin{bmatrix}1\\\bullet\ 1\end{bmatrix}\right) + R\alpha\left(P[101] + (z-2)\ P\begin{bmatrix}1\\0\ 1\end{bmatrix}\right)$$

$$\approx -R\alpha\ (z-1)\left(P[1]^2 - \frac{(P[1] - P[11])^2}{1 - P[1]}\right),$$

where the second term was approximated using PA, i.e. (4.16), and the first by totally ignoring correlations beyond NNs (i.e. $P[1\bullet1] = P[1]^2$). A similar approach was previously used in the context of a model for surface deposition by Filipe & Rodgers (1995). The positive triplet term ($\propto J$) in (4.57) was approximated as before using PA. Equation (4.59) can be rewritten as

$$-R\alpha\ (z-1)\left(\frac{P[11]^2}{P[1]} - P[1](1 - P[1])\ C(1)^2\right), \quad (4.60)$$

showing that by neglecting correlations more than in (4.16), (4.59) yields a less negative estimate of this triplet term, partially compensating for the underestimation of the positive triplet term in (4.57). It was verified that this approximation does indeed underestimate $P[1]$ (Filipe, 1999 a, b).

Having obtained two approximations for the negative triplet ($\propto R\alpha$) in (4.57) leading to opposite errors in the estimates of $P[1]$, the idea was to improve PA by using the "mixed" approximation

$$\delta \times [\text{ overestimating PA }] + (1 - \delta) \times [\text{ underestimating PA }], \quad (4.61)$$

where δ is a mixing parameter. $\delta = 1$ corresponds to the original PA and $\delta = 0$ to (4.59). This leads to

$$-R\alpha \left(P[111] + (z-2)\, P\begin{bmatrix} 1 & \\ 1 & 1 \end{bmatrix} \right) \tag{4.62}$$
$$\approx -R\alpha\,(z-1) \left[\frac{P[11]^2}{P[1]} - (1-\delta)\, P[1](1-P[1])\, C(1)^2 \right],$$

with the other triplet ($\propto J$) approximated as before. The next question was how to fix δ. Since the approximation given by (4.59) predicts that when $J_1 = 0$ the stationary state has a discontinuous phase transition, δ was chosen to be the smallest value for which the transition is continuous (Filipe, 1999 a). This gives

$$\delta = \frac{z(1+\lambda_c\alpha)[1+\lambda_c - z(1-\alpha)]}{\alpha(z-1)(z-\lambda_c)^2}, \tag{4.63}$$

where λ_c is the critical value of $\lambda = R/J$ (when $J_1 = 0$), given by

$$\lambda_c = z\left[-\epsilon + \sqrt{(\epsilon+1)^2 - 2(1+z\alpha)^2/(1+2z\alpha)}\right], \tag{4.64}$$

with $\epsilon = z[1+z\alpha(1+2\alpha)]/[2(1+2z\alpha)]$.

While HPA can only be devised for $\alpha > 0$, all results are continuous in the limit $\alpha \to 0$ and better than PA predictions. This is possible because $\delta\alpha \to -z(z-1-\lambda_c)/[(z-1)(z-\lambda_c)^2] \equiv \Delta$ as $\alpha \to 0$ (Filipe & Gibson, 2000). Therefore, the corresponding limit of the equation for $P[11]$ is not obtained by setting $\alpha = 0$ in (4.57) but by replacing the term given by (4.62) with the limit

$$-R\alpha \left(P[111] + (z-2)\, P\begin{bmatrix} 1 & \\ 1 & 1 \end{bmatrix} \right) \underset{\alpha\to 0}{\longrightarrow} -R\Delta\,(z-1)\, P[1](1-P[1])\, C(1)^2. \tag{4.65}$$

The HPA prediction for the persistence threshold of the Contact Process (i.e. $J_1 = 0$ and $\alpha = 0$) is given by

$$(J/R)_c = \frac{\sqrt{1+4\,(z-1)/z^2}+1}{2\,(z-1)}. \tag{4.66}$$

Table 4.1 lists the persistence threshold values of $J_2/R = zJ/R$, as predicted by all cluster approximations considered in this chapter and obtained from simulation for 1D and 2D systems. Figs 4.1 and 4.2 show the phase diagrams as predicted by all approximations and obtained from simulation. In 2D systems, HPA is definitely superior to the other approximations, especially in the proximity of the persistence threshold. Its improvement is not only quantitative (i.e.

providing the smallest mismatch between simulated and predicted threshold values) but also qualitative, since HPA is able to predict better the slope of the curve at the critical point and thus to capture the shape of the phase diagram. In 1D systems, HPA is still qualitatively better than all other approximations, predicting the correct shape of the simulated phase diagram, but it is only quantitatively superior to PA. The HPA threshold prediction coincides with that of 3A, but 3A provides better quantitative predictions for all other values of J_2. 4A is quantitatively superior to the other approximations over the whole parameter space.

	MF	PA	3A	4A	SA	HPA	simulation
1D	1	2	$2.41\cdots$	$2.63\cdots$	——	$2.41\cdots$	$3.24\cdots$
2D	1	$1.33\cdots$	——	——	$1.43\cdots$	$1.54\cdots$	$1.64\cdots$

Table 4.1: *Persistence threshold values of J_2 obtained from simulation and predicted by cluster approximations.*

4.3 Conclusions

The NN epidemic model is a particular case of the model introduced in chapter 2 with interactions restricted to the nearest neighbours of each infective on the lattice.

This chapter has focussed on the analytic description of the expected behaviour of that model. It reviewed some of the existing approaches with view to further developments in chapters 5 and 6. The dynamics of the model are described by an unlimited hierarchy of ordinary differential equations which can be solved approximately using closure assumptions to truncate the system. We reviewed existing cluster approximations and applied them to the NN epidemic model on a 1D linear lattice and on a 2D square lattice and compared their predictions to simulation.

Pairwise Approximation (PA) is the lowest–order improvement to the Mean Field approximation (MF), which corresponds to the non–spatial version of the model. By accounting for spatial correlations between neighbouring sites, PA describes the model behaviour significantly better than MF (Filipe & Gibson, 1998). Third and fourth–order cluster approximations (3A and 4A) to the model on a 1D lattice also yield an improvement relative to PA, as has already been shown for similar models in statistical physics (ben-Avraham & Köhler, 1992). When applied to 2D lattices, cluster approximations lead to systems of equations

which become progressively more complex as the order of the approximation increases. We considered a third–order approximation to the 2D lattice model which accounts for full spatial correlations between sites up to distance $\leq \sqrt{2}$ (SA) (Filipe & Gibson, 2000). SA, like 3A and 4A, performs better than PA. However, close to the persistence threshold (in the absence of background infection, i.e. $J_1 = 0$), all orders of approximation give qualitatively wrong predictions in one and two spatial dimensions. In the proximity of the threshold, in fact, the model develops long–range correlations (Filipe *et al.*, 2000) which these approximations, accounting for full correlations only within a finite basic cluster, are unable to capture. Cluster approximation assumptions imply that spatial correlations are always short–ranged; we showed, for example, that PA implies an exponential decay of correlations with the distance (Filipe, 2000, unpublished).

The inability of cluster approximations to predict the infinite slope of the phase diagrams at the persistence threshold affects predictions over a range of parameter values because epidemic curves become very steep near the threshold. An alternative approximation was recently proposed by Filipe (1999 a, b) in order to tackle this problem. The Hybrid Pairwise Approximation (HPA) is an elaboration of PA based on the idea of mixing two approximations, one overestimating and one underestimating the effects of local aggregation of infecteds. HPA was originally devised for a more general epidemic model of which the one considered here is a special case.

On 2D lattices, HPA substantially improves the other approximations over the whole range of parameters, both quantitatively and qualitatively, better predicting the location of the threshold and the slope of the stationary epidemic curve at that point. Although technically more complicated than PA, HPA is still simpler and more effective than higher–order cluster approximations.

On 1D lattices, HPA is superior to PA but not to higher–order cluster approximations. Both 3A and 4A give better quantitative predictions of the stationary epidemic size over the whole range of parameters considered. However, HPA is the approximation which better predicts the shape of the phase diagram.

Chapter 5

The anisotropic NN model

5.1 Introduction

Plant pathogens are transported from plant to plant by a large variety of agents, including wind, rain, soil, animals and humans. Whatever the vector, the mechanisms of propagule dispersal are influenced by a number of factors which might prevent a regular, isotropic spread of disease from an infective source. The topographical, vegetational and meteorological characteristics of the environment in which the epidemic develops affect the way in which propagules are transported and deposited and consequently influence the progress of the disease.

Clues about the mode of spread of a disease often come from the observation of its spatial patterns and the analysis of the shape of the infected regions (Van de Lande & Zadoks, 1999). A preferential direction, such as along rows, could indicate mechanical transmission through harvesting implements; on the other hand, a directional distribution associated with the prevalent wind might suggest airborne dispersal. The effects of wind on plant disease dispersal have been the subject of a large number of experimental studies (Gregory, 1973; Aylor, 1987; Fitt & McCartney, 1986; Zawolek, 1993). Other reasons for the anisotropic spread of plant epidemics are the heterogeneity of the soil and the topography of the territory. Soil–borne diseases, such as the "red core root disease" of strawberries, are closely related to high soil moisture content. Their spread in fields with inclinations which favour the flow of water in a particular direction can be highly anisotropic (Hickman, 1940). The transmission of disease along a preferential direction may also occur as an effect of rectangular rather than square plots, where the distances between rows and columns are different (Aylor & Ferrandino, 1989, 1990). This latter situation is in fact very common in experimental fields.

Despite the notable amount of experimental evidence suggesting that plant diseases spread anisotropically in space, epidemic models typically assume that transmission occurs evenly in all directions. In this chapter we are concerned

with a generalisation of the NN model (introduced in the previous chapter) which assumes anisotropic spread of the disease: the infection would still spread from an infected to its four neighbouring sites but with different strengths depending on the direction.

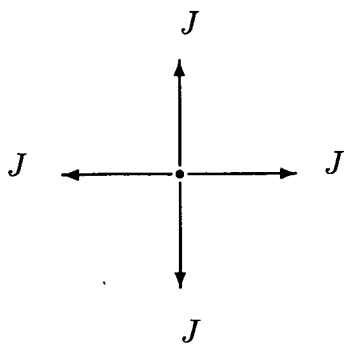
First of all we formulate the model and investigate its behaviour through computer simulation. We then derive a system of differential equations which describe the dynamics of the model and form an infinite hierarchy of equations, as in the case of the isotropic NN model. The closure techniques introduced in chapter 4 are extended to this system of equations and approximate solutions are obtained and compared to simulation. In this way, we assess whether the closure approximations considered can capture the effects of anisotropy on disease spread.

5.2 Spatial anisotropy

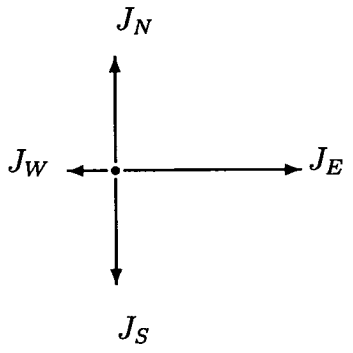
In the NN epidemic model introduced in the previous chapter, J_2 represents the rate at which each infected produces pathogens which are spread evenly in all directions. The rate of deposition on each NN is given by $J = J_2/z$ and is constant. We include spatial anisotropy in the model by keeping the sporulation rate per infected, J_2, constant but making the rate of deposition of propagules J dependent on the direction of spread. We restrict our attention to a square lattice ($z = 4$) and consider different rates of deposition of spores along the horizontal and vertical axes and, within the horizontal axis, different rates of deposition towards East (E) and West (W) (or analogously, within the vertical direction, different towards North (N) and South (S)). To summarise, there are four possible directions for the deposition of propagules and two constraints: 1) the total deposition rate of spores dispersed in all directions must equal the sporulation rate J_2; 2) along the vertical (or horizontal) axis the deposition rates of spores towards North and South (or East and West) must be equal. Hence two new parameters are introduced, $\rho_E, \rho_W \in [-1, 1]$, and the deposition rates are defined as:

- towards East: $J_E = J(1 + \rho_E)$
- towards West: $J_W = J(1 + \rho_W)$
- towards North: $J_N = J[1 - (\rho_E + \rho_W)/2]$
- towards South: $J_S = J[1 - (\rho_E + \rho_W)/2]$.

Graphically the situation, for $\rho_E > 0$ and $\rho_W < 0$, is the following:



isotropic model



anisotropic model

A more general model with four (rather than three) different deposition rates could be formulated (by removing constraint 2) above) but for reasons of simplicity we shall restrict our attention to the case described.

We now derive dynamic equations for the densities of infecteds and of pairs of infecteds corresponding to the anisotropic model. The basic difference relative to equations (4.1) and (4.2) is that we must distinguish between vertical and horizontal pairs, since in general we expect that $P[11] \neq P\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$, and similarly $P[111] \neq P\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$. If, for example, $\rho_E = 0$ and $\rho_W = 1$, infected regions will be elongated horizontally, and $P[11] > P\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$. In addition, since $P[10] = P[1] - P[11]$ and $P\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right] = P[1] - P\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$, we also have $P[10] < P\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$. The same relations apply to $P[01]$ and $P\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$, hence we still have $P[10] = P[01]$, $P\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right] = P\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$, $P\left[\begin{smallmatrix}1&0\end{smallmatrix}\right] = P\left[\begin{smallmatrix}0&1\end{smallmatrix}\right]$, and $P\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right] = P\left[\begin{smallmatrix}1&0\end{smallmatrix}\right]$.

The equations can be derived in a similar way to the isotropic case, giving

$$\frac{dP[1]}{dt} = J_1 - P[1]\,(J_1 - 4J + R) - P[11]\,J(2+\rho_h) - P\begin{bmatrix}1\\1\end{bmatrix}\,J(2-\rho_h) \quad (5.1)$$

$$\begin{aligned}\frac{dP[11]}{dt} &= P[1]\,[2J_1 + J(2+\rho_h)] - P[11]\,[2J_1 + J(2+\rho_h) + 2R] \\ &+ P[101]\,J(2+\rho_h) + \left(P\begin{bmatrix}1\\0&1\end{bmatrix} + P\begin{bmatrix}&1\\1&0\end{bmatrix}\right) J(2-\rho_h) \end{aligned} \quad (5.2)$$

$$\begin{aligned}\frac{dP\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]}{dt} &= P[1]\,[2J_1 + J(2-\rho_h)] - P\begin{bmatrix}1\\1\end{bmatrix}\,[2J_1 + J(2-\rho_h) + 2R] \\ &+ P\begin{bmatrix}1\\0\\1\end{bmatrix}\,J(2-\rho_h) + \left(P\begin{bmatrix}1\\0&1\end{bmatrix} + P\begin{bmatrix}&1\\1&0\end{bmatrix}\right) J(2+\rho_h), \end{aligned} \quad (5.3)$$

where $\rho_h = \rho_E + \rho_W$. The system of equations (5.1)–(5.3) reduces to (4.1)–(4.2) when $\rho_E = \rho_W = 0$.

In particular, if $\rho_E = \rho_W \equiv \rho$, there are only two different contact rates, along the vertical and horizontal axes. This particular model describes the kind of anisotropy which occurs in fields where the distances between rows and columns are different. In this case, strictly speaking, if d_r is the distance between rows, d_c is the distance between columns and $d_r < d_c$, the anisotropic NN model can only describe situations where $d_c < 2d_r$. To model the case $d_c \geq 2d_r$ we should take into account interactions between individuals on the same row at distances larger than 1 lattice unit, at least up to d_c, in which case the model would no longer be a NN model.

In this study we have focussed on the case $\rho = \rho_E = \rho_W$, and this is the model to which we shall refer in the rest of the chapter.

5.3 Simulation results

First the model behaviour is investigated using simulation. We consider a square lattice of linear size $L = 100$ with toroidal periodic boundary conditions. The epidemics are initiated with 0.2% of the hosts randomly selected to be infected. Anisotropic effects are expected to be more evident in the absence of background infection ($J_1 = 0$), which is a random, non–spatial process; hence we concentrate on the case $J_1 = 0$. All the simulation results shown in this section are averaged over five stochastic realisations and conditional on epidemic survival.

We study the epidemic behaviour over a range of values of ρ. When $\rho = 0$, the isotropic NN model is recovered; as ρ increases, the model becomes more and more anisotropic, favouring horizontal spread of the disease. As a result, elongated domains of infection are observed (see Fig. 5.8). When $\rho = 1$, the model is completely anisotropic: the disease can only spread within each horizontal row. Analogously, as ρ decreases from 0 to -1, spread increases along the vertical direction. Since the two situations are completely symmetrical, we shall restrict our attention to the first case. When $\rho = 1$, if the initial number of randomly distributed infectives is small and there is no background infection, some horizontal rows will contain only healthy individuals. Since the disease cannot spread across rows, these rows will remain free from the disease throughout the process. Extreme situations, such as $\rho = 1$ and $\rho = -1$, where dispersal in one direction is completely forbidden, may seem unrealistic. Yet experimental situations can exist where some sort of "barrier" is created to avoid between–row interactions (e.g. in an experiment to investigate the Lotka–Volterra predator–prey model,

Huffaker (1958) used barriers of vaseline to prevent migration of predators over specific boundaries (Renshaw, 1991)). This limiting case is also interesting from a mathematical point of view as it represents a transition between 2D and 1D behaviour.

For a given value of the sporulation rate J_2, we examine the transient and stationary behaviours as ρ varies from 0 to 1. There are two possible choices for the value of J_2, namely lower or higher than the persistence threshold of the 1D isotropic model. Fig. 5.1 shows the phase diagrams of the (isotropic) NN model corresponding to 1 and 2 spatial dimensions. For a given value of J_2, the stationary epidemic size in the 2D system is higher than in the 1D system and the threshold value is smaller in 2D than in 1D. This is intuitively easy to understand:

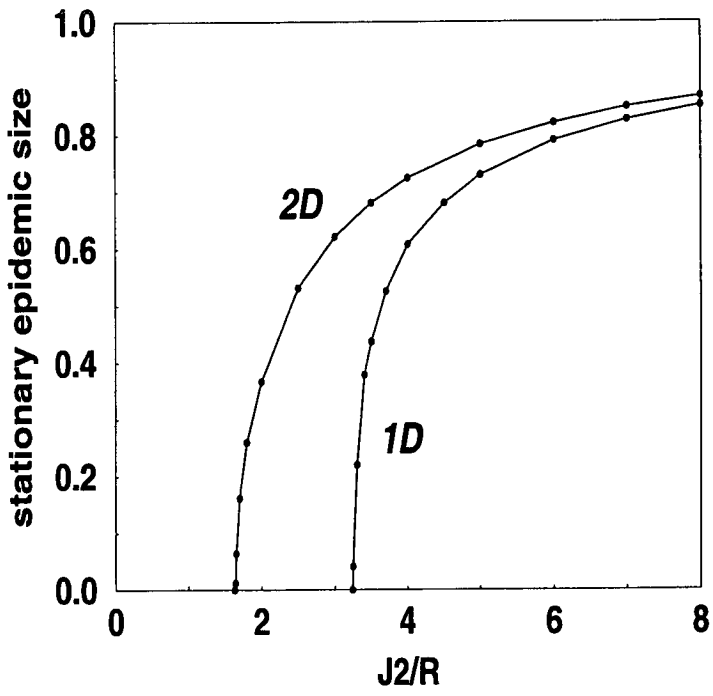


Figure 5.1: *Phase diagrams for epidemic models in 1 and 2 spatial dimensions obtained from simulation.*

whilst on a 1D lattice there is only one possible "path" between a given susceptible and an infected, on a 2D lattice there are many, and hence a larger level of infection is expected for a given value of J_2. We expect that by varying ρ from 0 to 1 the equilibrium size of the 2D epidemic will gradually decrease. If J_2 is chosen between the two isotropic thresholds (i.e. $(J_2^{1D})_c \geq J_2 \geq (J_2^{2D})_c$, case I), then as ρ increases, transmission of the disease across rows becomes increasingly unlikely; and since J_2 is not large enough to sustain an epidemic on a 1D lattice, the epidemic eventually dies out. In this case, epidemics always die out before the value $\rho = 1$ is reached. On the other hand, if J_2 is chosen above the 1D

threshold (i.e. $J_2 > (J_2^{1D})_c$, case II), the epidemic has non–zero stationary levels for all values of ρ. We shall consider case I and case II separately.

5.3.1 Case I: $(J_2^{1D})_c \geq J_2 \geq (J_2^{2D})_c$

Fig. 5.2 illustrates the time evolution of the epidemic size for $J_2 = 2$ and values of ρ ranging from 0 to $\rho_c = 0.77$. When $\rho = \rho_c$ the epidemic dies out. As

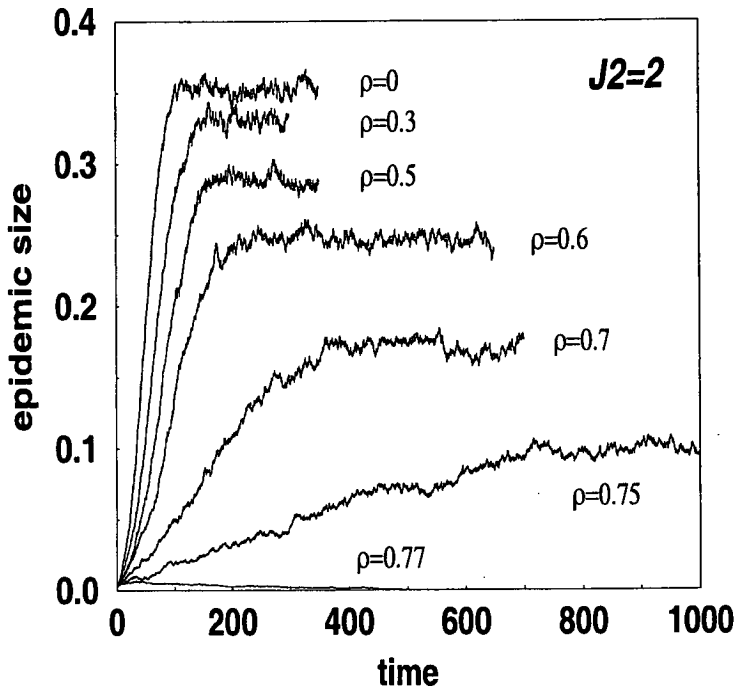


Figure 5.2: *Case I. Epidemic size vs time; $J_2 = 2$, ρ varying from 0 (isotropic model) to $\rho_c = 0.77$, at which the epidemic becomes extinct.*

expected, the stationary epidemic size decreases and the transient length increases as the anisotropy parameter ρ increases. By plotting the equilibrium epidemic size against ρ for a given J_2, we obtain a "transversal" phase diagram, which is shown in Fig. 5.3. This exhibits a sharp continuous transition from epidemic persistence to extinction. Fig. 5.4 shows the time evolution of $P[11]$ and $P\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$ (indicated as $P[11]_{horizontal}$ and $P[11]_{vertical}$, respectively, in the figure) for some values of ρ. At any given time t, both densities decrease as ρ increases; for a given $\rho > 0$, $P\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right](t) < P[11](t)$. Fig. 5.5 shows their relative difference $([P[11]-P\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]]/P[11])$ at equilibrium.

5.3.2 Case II: $J_2 > (J_2^{1D})_c$

Case II is illustrated in Fig. 5.6. We choose $J_2 = 6$ and observe the time evolution of $P[1]$, $P[11]$ and $P\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$. The effects of anisotropy are very different in this case.

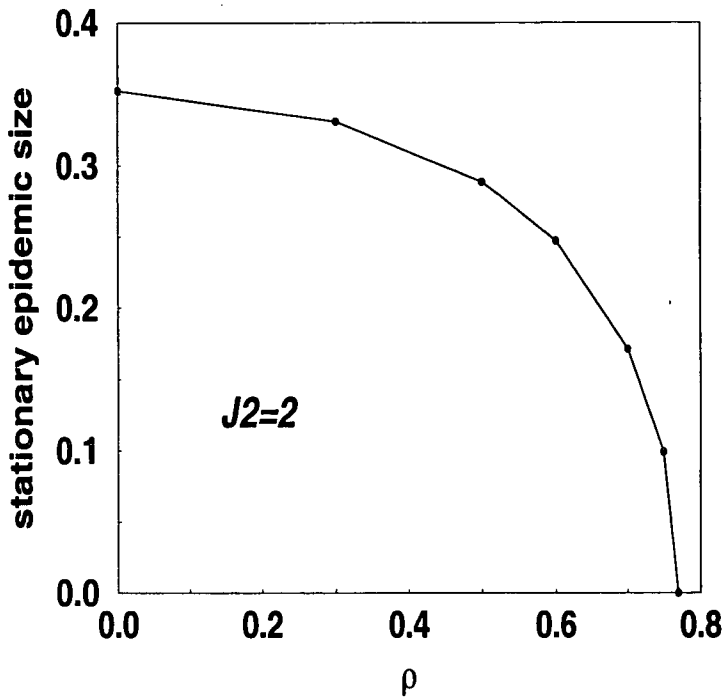


Figure 5.3: *Case I. Stationary epidemic size at fixed $J_2 = 2$ vs different values of the anisotropy parameter ρ.*

The equilibrium level of infection is high and hardly varies as ρ increases. For $\rho = 0.9999$, $P[1]_\infty$ is only $\sim 5\%$ smaller than the isotropic value (not visible in Fig. 5.6). Only when $\rho = 1$, i.e. in the fully anisotropic case where epidemics develop independently in different rows, does the equilibrium epidemic size drop dramatically. We note that this value is not the same as that in the 1D limit of the model, which is considerably higher and closer to the 2D isotropic value (Fig. 5.1). The equilibrium disease level would be the same as the 1D isotropic model only if we measured the infection level in each separate row and conditioned the measures on epidemic survival rather than considering the whole system. The epidemics are initiated by a small number of randomly distributed infecteds (0.2% of the hosts): this implies that some rows are completely susceptible at the start of the epidemic, and since the disease cannot be transmitted from one row to another, they will never become infected throughout the process. In addition, in some rows the epidemic might become extinct (note that in this example each row consists of only 100 individuals). If we exclude the rows where the epidemic never started or died out, we recover the 1D value for the epidemic size at equilibrium ($\sim 79\%$, see Fig. 5.1).

In intermediate situations ($0 < \rho < 1$), the epidemics grow to an asymptotic value between the two isotropic limits ($P[1]_\infty \simeq 82\%$ for 2D and $P[1]_\infty \simeq 79\%$ for 1D). However, transients become extremely slow, especially for ρ near 1. The

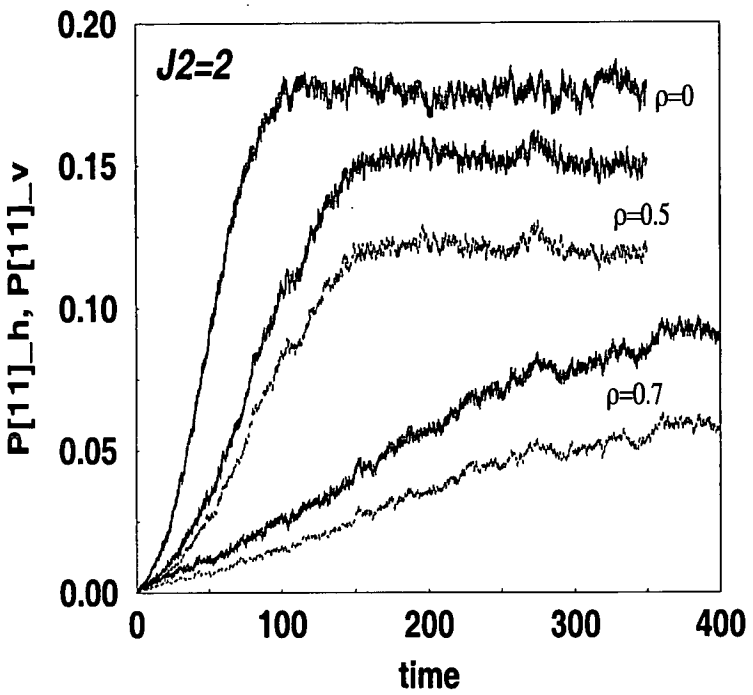


Figure 5.4: *Case I. Time evolution of* $P[11]_{horizontal}$ *(thick solid line) and* $P[11]_{vertical}$ *(thin broken line) for some values of* ρ *(note that for* $\rho = 0$ *the two curves overlap).*

same effect is also observed in case I ($J_2 = 2$) but it is much weaker. This results from the fact that, although disease spreads very rapidly within rows which are initially infected (because J_2 is large), it might take a long time before the disease is transmitted to rows which begin without infectives. This also explains why for $\rho = 1$ the transient is comparatively short: the stationary epidemic size in this case corresponds to the stationary epidemic size of separate infected rows.

Obviously, different sets of initial conditions would lead to completely different situations. For example, if the epidemic starts with one infected in each row and

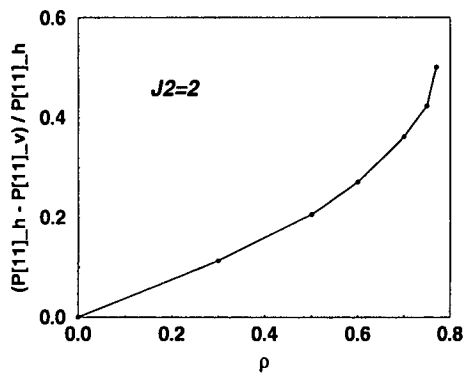


Figure 5.5: *Case I. Relative difference between* $P[11]_{horizontal}$ *and* $P[11]_{vertical}$ *at equilibrium for some values of* ρ.

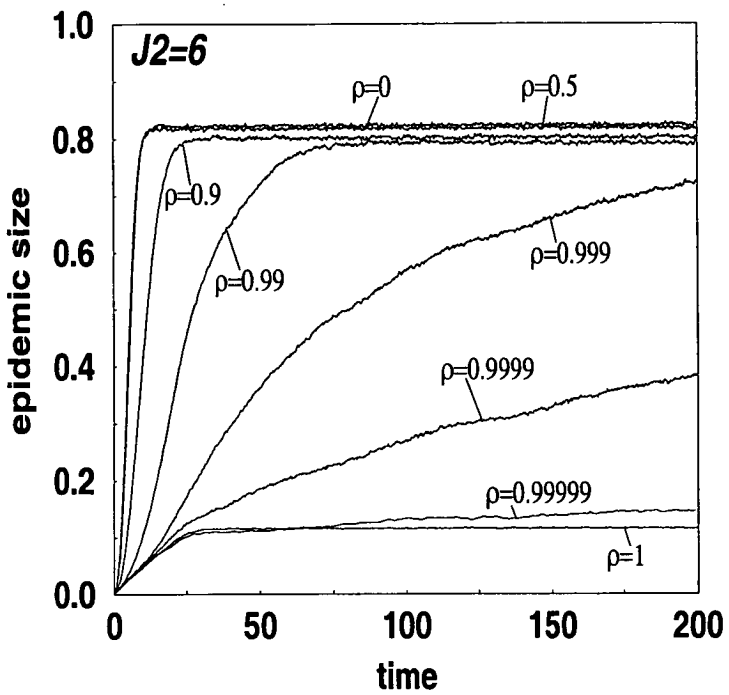


Figure 5.6: *Case II. Epidemic size vs time; $J_2 = 6$, ρ varying from* 0 *(isotropic model) to* 1 *(totally anisotropic model).*

the measure of the stationary epidemic size is conditioned on epidemic survival in each row, then the behaviour of the 1D model is recovered.

The transversal phase diagram (Fig. 5.7) for $J_2 = 6$ is therefore very different from the one for $J_2 = 2$ (Fig. 5.3). The epidemic size is almost constant for most values of ρ; the sharp, discontinuous transition at $\rho = 1$ reflects the initial conditions chosen.

5.3.3 Comments

Case II does not have great practical relevance since it deals with large long–term epidemics where spatial effects are not very important and which can be described reasonably well by non–spatial models. The effects of anisotropy on the stationary behaviour are irrelevant, except in the limit $\rho = 1$, when the system is completely anisotropic and the stationary behaviour depends on the initial condition. $\rho = 1$ is a limit case which is unlikely to occur in real systems. The main effect of the anisotropic spread on the transient behaviour is to increase the transient duration.

Case I is more interesting since it describes smaller stationary epidemics where spatial aspects of disease transmission and, in particular, anisotropic effects are more important. As the spread of the disease becomes more anisotropic, the stationary size of the epidemic gradually decreases and the transient length increases. For a given J_2, there is a threshold value of the anisotropic parameter ρ

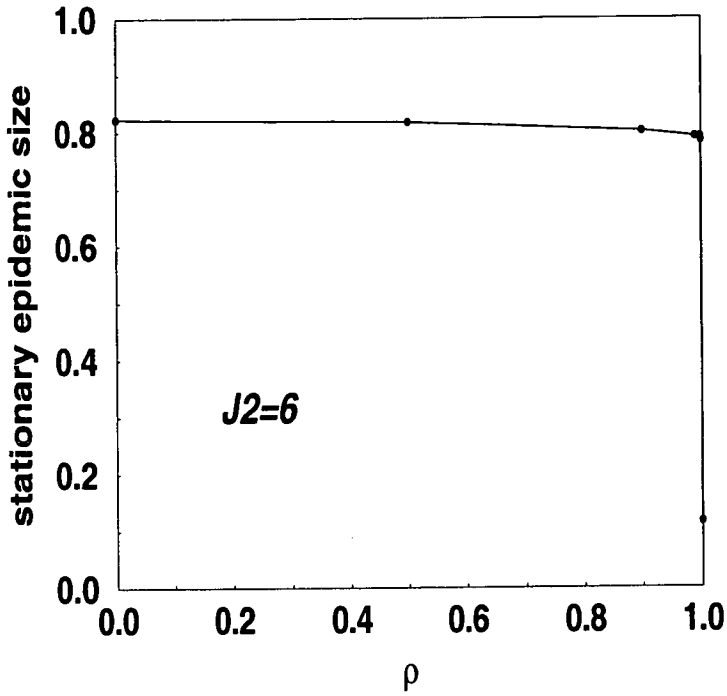


Figure 5.7: *Case II. Stationary epidemic size at fixed $J_2 = 6$ vs different values of the anisotropy parameter ρ.*

separating regions of epidemic persistence and extinction.

Fig. 5.8 shows three snapshots of the disease at equilibrium with similar epidemic sizes ($\sim 10\%$): in the isotropic case ($J_2 = 1.688, \rho = 0$), in an intermediate situation, where domains of infections start to be elongated ($J_2 = 2, \rho = 0.75$), and in the fully anisotropic case, where the epidemic develops separately in different rows ($J_2 = 6, \rho = 1$).

5.4 Approximations

We now use closure approximations to close the system of equations (5.1)–(5.3) and solve it numerically. We wish to compare the predictions of the various approximations introduced in chapter 4 to simulation of the anisotropic NN model. The model we simulated corresponds to a particular case of (5.1)–(5.3) where $\rho_E = \rho_W$ and $P\left[\begin{smallmatrix}1\\0\end{smallmatrix}\begin{smallmatrix}\\1\end{smallmatrix}\right] = P\left[\begin{smallmatrix}\\1\end{smallmatrix}\begin{smallmatrix}1\\0\end{smallmatrix}\right]$. The MF approximation performs very poorly even in the isotropic case. Since MF assumes that $P[11] = P\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right] = P[1]^2$, it cannot capture the effects of anisotropy at all.

Within PA (Filipe & Gibson, 1998) we close the system (5.1)–(5.3) assuming the following relations:

$$P[101] \approx P[10]^2/P[0] = (P[1] - P[11])^2/(1 - P[1]) \tag{5.4}$$

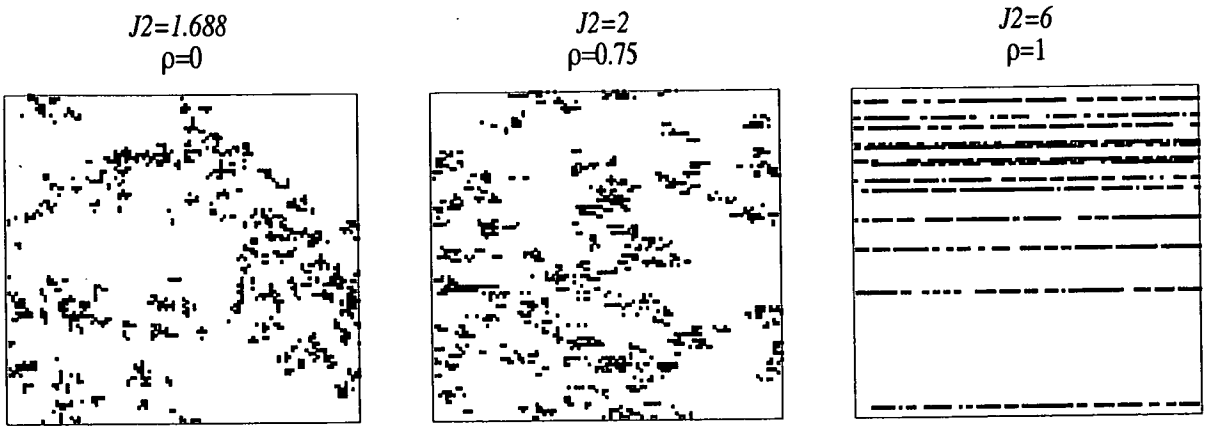


Figure 5.8: *Snapshots of three epidemics with stationary epidemic size* $\sim 10\%$ *in a* 100×100 *isotropic system (left), a system with* $\rho = 0.75$ *(centre), and a completely anisotropic system (right).*

$$P\begin{bmatrix}1\\0\\1\end{bmatrix} \approx P\begin{bmatrix}1\\0\end{bmatrix}^2 / P[0] = \left(P[1] - P\begin{bmatrix}1\\1\end{bmatrix}\right)^2 / (1 - P[1]) \tag{5.5}$$

$$P\begin{bmatrix}1&\\0&1\end{bmatrix} = P\begin{bmatrix}&1\\1&0\end{bmatrix} \approx P[10]\ P\begin{bmatrix}1\\0\end{bmatrix} / P[0] =$$
$$(P[1] - P[11])\left(P[1] - P\begin{bmatrix}1\\1\end{bmatrix}\right) / (1 - P[1]). \tag{5.6}$$

Applying SA (Filipe & Gibson, 2000) to the most general case, where $\rho_E \neq \rho_W$, means solving 8 equations, because the densities of differently oriented triplets are themselves different. Restricting our attention to the current case, where $\rho = \rho_E = \rho_W$, reduces the number of the equations to 6: equations (5.1)–(5.3), and

$$\begin{aligned}\frac{dP\left[\begin{smallmatrix}1\\0&1\end{smallmatrix}\right]}{dt} = & \quad J_1\left(P\begin{bmatrix}0&\\0&1\end{bmatrix} + P\begin{bmatrix}1&\\0&0\end{bmatrix} - P\begin{bmatrix}1&\\0&1\end{bmatrix}\right) \\ + & \quad R\left(P\begin{bmatrix}1&\\1&1\end{bmatrix} - 2P\begin{bmatrix}1&\\0&1\end{bmatrix}\right) - 2J\ P\begin{bmatrix}1&\\0&1\end{bmatrix} \\ + & \quad J(1+\rho)\left(P\begin{bmatrix}1&0\\&0&1\end{bmatrix} + P\begin{bmatrix}0&1\\0&1\end{bmatrix} + P\begin{bmatrix}1&&\\0&0&1\end{bmatrix} - P\begin{bmatrix}&1&\\1&0&1\end{bmatrix}\right) \\ + & \quad J(1-\rho)\left(P\begin{bmatrix}1&\\0&\\0&1\end{bmatrix} + P\begin{bmatrix}1&1\\0&0\end{bmatrix} + P\begin{bmatrix}1&0\\&0\end{bmatrix} - P\begin{bmatrix}1&0&1\\1\end{bmatrix}\right)\end{aligned} \tag{5.7}$$

$$\frac{dP\left[\begin{smallmatrix}1\\1\,1\end{smallmatrix}\right]}{dt} = J_1\left(P\begin{bmatrix}0 & \\1 & 1\end{bmatrix} + P\begin{bmatrix}1 & \\1 & 0\end{bmatrix} - P\begin{bmatrix}1 & \\0 & 1\end{bmatrix}\right) \tag{5.8}$$
$$- \; 3R\,P\begin{bmatrix}1 & \\1 & 1\end{bmatrix} + 2J\,P\begin{bmatrix}1 & \\0 & 1\end{bmatrix} + J(1+\rho)\,P\begin{bmatrix}1 & \\1 & 0\end{bmatrix} + J(1-\rho)\,P\begin{bmatrix}0 & \\1 & 1\end{bmatrix}$$
$$+ \quad J(1+\rho)\left(P\begin{bmatrix}1 & & \\1 & 0 & 1\end{bmatrix} + P\begin{bmatrix} & 1 & \\1 & 0 & 1\end{bmatrix} + P\begin{bmatrix}0 & 1 & \\1 & 1 & \end{bmatrix} + P\begin{bmatrix}1 & 0 & \\ & 1 & 1\end{bmatrix}\right)$$
$$+ \quad J(1-\rho)\left(-P\begin{bmatrix}1 & \\0 & \\1 & 1\end{bmatrix} + P\begin{bmatrix}1 & \\0 & 1\\1 & \end{bmatrix} + P\begin{bmatrix}1 & 1\\1 & 0\end{bmatrix} - P\begin{bmatrix}1 & \\1 & 0\\ & 1\end{bmatrix}\right)$$

$$\frac{1}{4}\frac{dP\left[\begin{smallmatrix}1\,1\\1\,1\end{smallmatrix}\right]}{dt} = (J_1+2J)\,P\begin{bmatrix}0 & 1\\1 & 1\end{bmatrix} - R\,P\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix} \tag{5.9}$$
$$+ \; J(1+\rho)\,P\begin{bmatrix}1 & 0 & 1\\ & 1 & 1\end{bmatrix} + J(1-\rho)\,P\begin{bmatrix}1 & \\0 & 1\\1 & 1\end{bmatrix}.$$

The system of equations is closed using the assumption of conditional independence between sites at distance larger than $\sqrt{2}$, as done for the isotropic model (see (4.49)–(4.55); Filipe & Gibson, 2000). Clusters with different orientations must now be decoupled in different ways; for example now we have

$$P[101] \approx \frac{P[10]^2}{P[0]} \neq P\begin{bmatrix}1\\0\\1\end{bmatrix} \approx \frac{P\begin{bmatrix}1\\0\end{bmatrix}^2}{P[0]} \tag{5.10}$$

and

$$P\begin{bmatrix}1 & \\0 & 0 & 1\end{bmatrix} \approx \frac{P\begin{bmatrix}1 & \\0 & 0\end{bmatrix}\,P[10]}{P[0]} \neq P\begin{bmatrix}1 & \\0 & \\0 & 1\end{bmatrix} \approx \frac{P\begin{bmatrix}0 & \\0 & 1\end{bmatrix}\,P\begin{bmatrix}1\\0\end{bmatrix}}{P[0]}, \tag{5.11}$$

and so on. Similarly, the exact relations used in chapter 4 now depend on the orientation of the clusters; for example

$$P\begin{bmatrix}1 & \\1 & 0\end{bmatrix} = P\begin{bmatrix}1\\1\end{bmatrix} - P\begin{bmatrix}1 & \\1 & 1\end{bmatrix} \neq P\begin{bmatrix}0 & \\1 & 1\end{bmatrix} = P[11] - P\begin{bmatrix}1 & \\1 & 1\end{bmatrix}, \tag{5.12}$$

and so on.

HPA (Filipe, 1999 a, b) is applied to (5.2) and (5.3) in a way that is analogous to the isotropic case. First we consider a more general model where infected neighbours of recovered individuals recover with probability α. HPA is obtained by using a linear combination of overestimating (PA) and underestimating approximations to the terms proportional to α in the equations for $dP[11]/dt$ and

$dP\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]/dt$ (see (4.58) and (4.59)). These terms are:

$$P[111]+2P\left[\begin{smallmatrix}1\\1\,1\end{smallmatrix}\right]=\begin{cases}\dfrac{P[11]^2}{P[1]}+2\dfrac{P[11]P\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]}{P[1]} & \text{overestimating}\\ 3P[1]^2-\dfrac{P[10]^2}{P[0]}-2\dfrac{P[10]P\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]}{P[0]} & \text{underestimating}\end{cases} \tag{5.13}$$

and

$$P\left[\begin{smallmatrix}1\\1\\1\end{smallmatrix}\right]+2P\left[\begin{smallmatrix}1\\1\,1\end{smallmatrix}\right]=\begin{cases}\dfrac{P\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]^2}{P[1]}+2\dfrac{P[11]P\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]}{P[1]} & \text{overestimating}\\ 3P[1]^2-\dfrac{P\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]^2}{P[0]}-2\dfrac{P[10]P\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]}{P[0]} & \text{underestimating}\end{cases} \tag{5.14}$$

We then consider the limit for $\alpha \to 0$ and obtain a version of HPA which we can apply to our anisotropic NN model.

Disease progress curves predicted by the above approximations, for cases II and I (section 5.3), are shown and compared to simulation in Figs 5.9 and 5.10, respectively. These results are discussed in the next sections.

5.4.1 Case II: $J_2 = 6$

For $J_2 = 6$, all approximations give very accurate predictions for the stationary epidemic size. Disease progress curves predicted by the spatial approximations (PA, SA and HPA) are almost indistinguishable at equilibrium and very close to the simulation curve for $\rho \leq 0.99$. Only MF, which is non–spatial and isotropic, slightly overestimates the asymptotic disease level (Fig. 5.9). Large values of the sporulation rate, such as $J_2 = 6$, generate epidemics in which spatial effects, and in particular anisotropy, are not very important. Hence, even non–spatial approximations such as MF are able to describe some aspects of the epidemics quite accurately.

None of the approximations are able to predict the rise in the duration of transients due to the increase in anisotropy. MF predicts the worst (that is, shortest) and HPA the best (i.e. longest) transient, although the difference is very small. This inability of the approximations to predict epidemic transients is not unexpected. The transient duration depends not only on the initial number of infecteds but also on their spatial distribution (see section 5.3); since only limited information is contained in the dynamic equations, this aspect of the model behaviour is not well captured by any approximation.

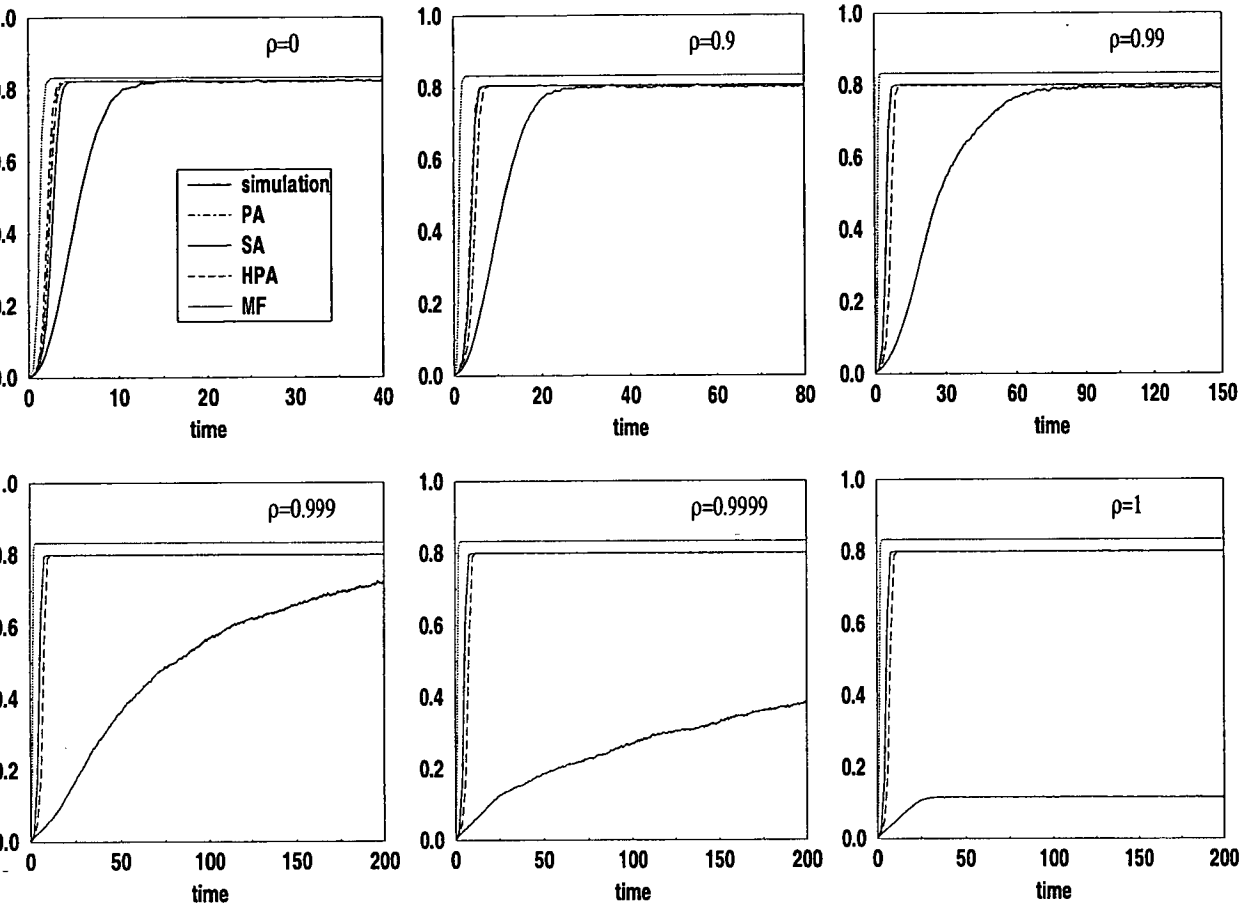


Figure 5.9: *Case II. Epidemic size vs time. Comparison of disease progress curves obtained from simulation and predicted by cluster approximations.* $J_2 = 6$, ρ *varying from* 0 *(isotropic model) to* 1 *(totally anisotropic model). Approximation curves from left to right (worst to best):* $\rho = 0$*: MF, PA, HPA, SA;* $\rho > 0$*: MF, PA and SA (indistinguishable), HPA.*

5.4.2 Case I: $J_2 = 2$

For $J_2 = 2$, all approximations overestimate asymptotic disease levels, even for the isotropic model. As the anisotropic effects increase ($\rho \to 1$), this overestimation becomes larger. The predicted stationary epidemic size does decrease as ρ increases (with the exception of MF, which in fact does not depend on the value of ρ), but not as much as in the simulation. In particular, the approximations fail to predict the extinction of the epidemic at $\rho_c = 0.77$ (Fig. 5.10).

SA performs better than PA, although the difference between the two predictions is quite small. HPA still overestimates the stationary epidemic size and underestimates the transient length but its predictions are undoubtably the most accurate. In the isotropic case (first graph of Fig. 5.10), the HPA prediction for

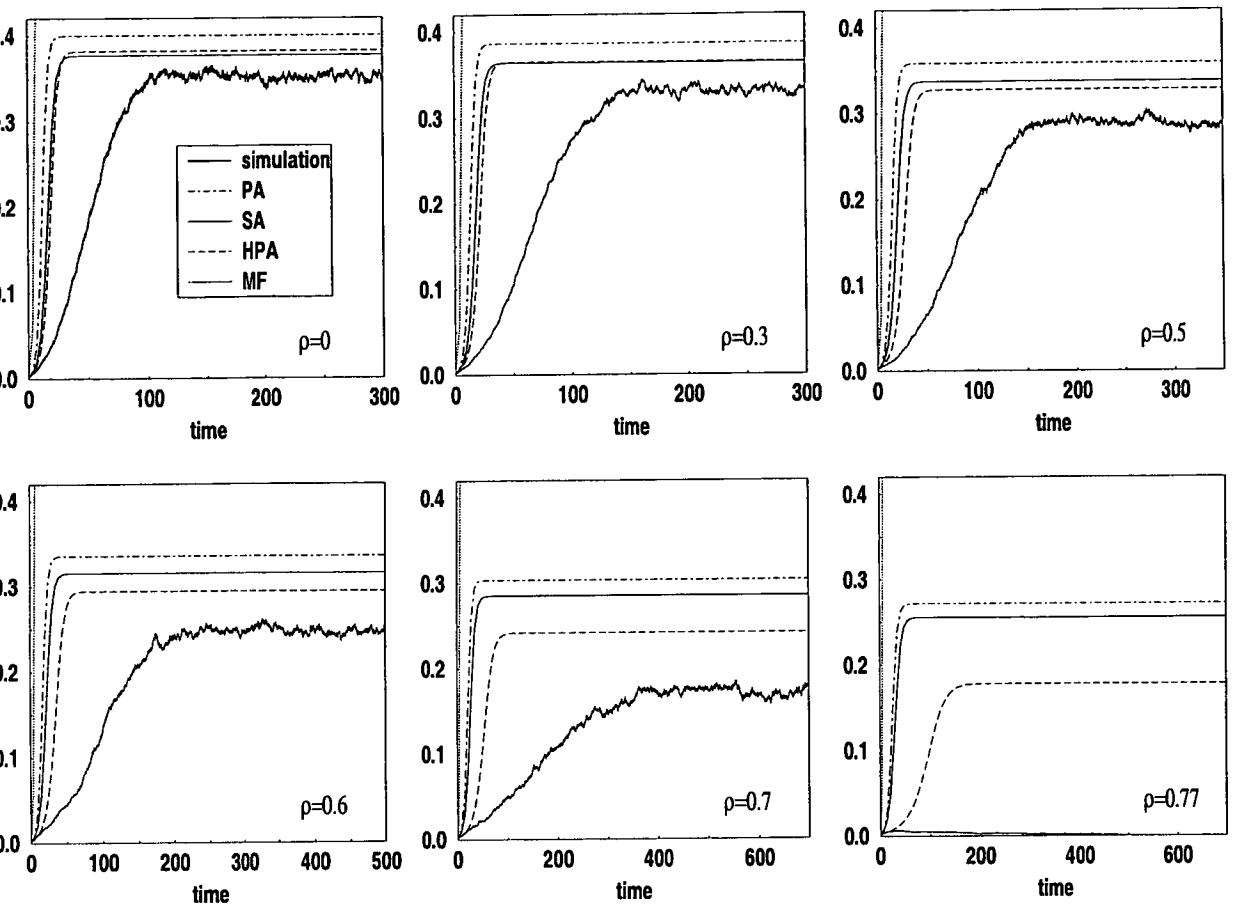


Figure 5.10: *Case I. Epidemic size vs time. Comparison of disease progress curves obtained from simulation and predicted by cluster approximations.* $J_2 = 2$, ρ *varying from* 0 *(isotropic model) to* $\rho_c = 0.77$, *at which the epidemic becomes extinct. Approximation curves from left to right (worst to best): MF, PA, SA, HPA.*

the asymptotic disease level is slightly above the SA prediction but the difference between the two is almost negligible (see also Fig. 5.11).

Figs 5.11 and 5.12 show how the equilibrium epidemic size varies with ρ for fixed J_2 and with J_2 for fixed ρ, respectively. The approximated phase diagrams are compared to simulation. HPA is superior to the other approximations in predicting the correct shape and the infinite slope of the curve in the proximity of the persistence threshold. Among all the approximations, HPA is the most sensitive to the spatial anisotropy of the model. For small values of ρ ($\sim 0 \leq \rho \leq 0.3$) HPA remains slightly higher than SA - which explains why in Fig. 5.10, for $\rho = 0$, SA predicts the equilibrium epidemic size better than HPA - although the difference between the two approximations is very small and in no way affects the merits of HPA.

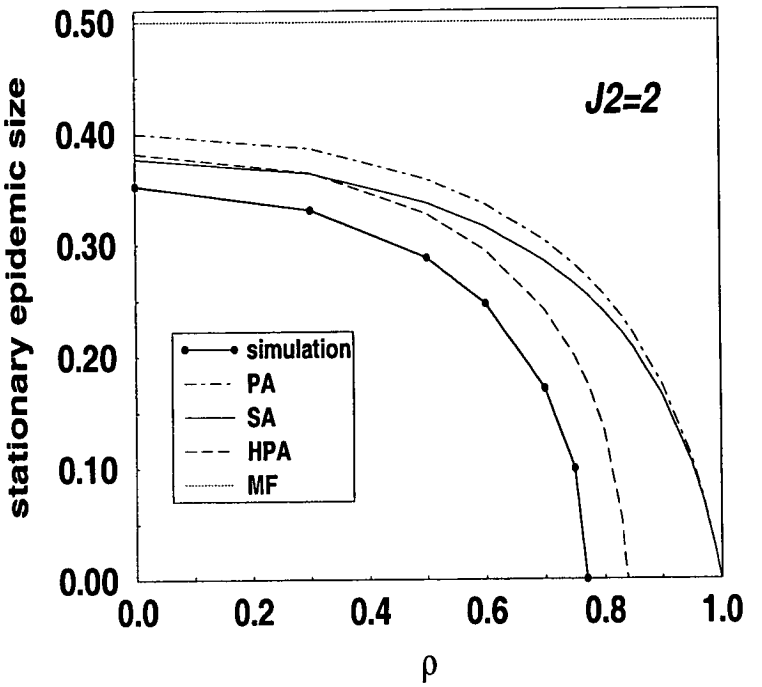


Figure 5.11: *Case I. Stationary epidemic size at fixed $J_2 = 2$ vs different values of the anisotropy parameter ρ as obtained from simulation and predicted by the approximations.*

5.4.3 Case I. Model with background infection

One way of making the model more realistic is to include some form of background infection (Gibson, 1997; Filipe & Gibson, 1998). In real systems it is unlikely that infectious contacts occur exclusively among NNs. Moreover, fields or plantations are never completely isolated and it is plausible that some spores reach the system from external sources of infection.

A background infection with rate 200 times smaller than the total sporulation rate ($J_1 = 0.01, J_2 = 2$) was included in the model and, once again, approximations were compared to simulation (Fig. 5.13). In this case the epidemic does not die out even in the 1D limit ($\rho = 1$).

As expected, MF performance is again extremely poor (Fig. 5.13).

PA predictions are better but increasingly overestimate the asymptotic behaviour as anisotropy increases. SA performs slightly better than PA, except for $\rho = 1$ where PA and SA curves are indistinguishable.

Very good agreement is found between the simulation disease levels at equilibrium and HPA predictions. HPA is able to capture the effects of anisotropy up to the limit value $\rho = 1$.

As before, the approximations are inaccurate in predicting transient behaviour.

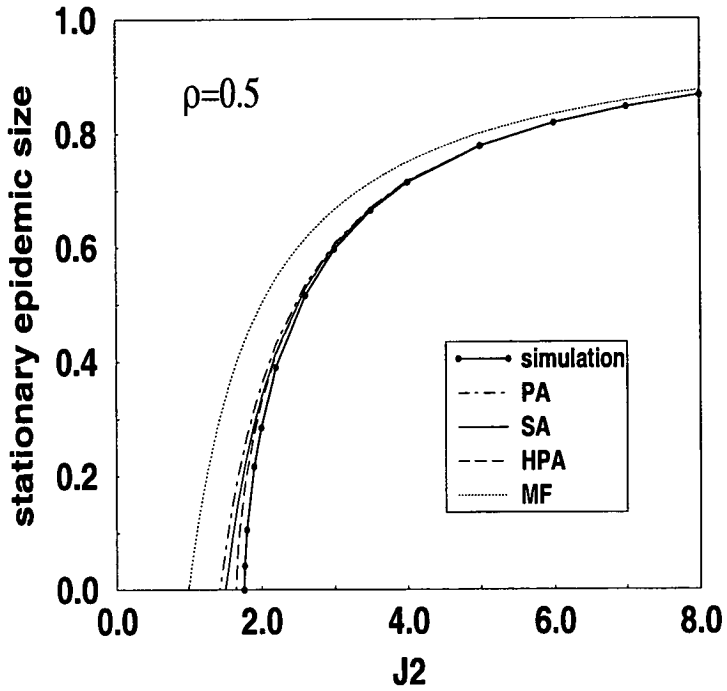


Figure 5.12: *Case I. Stationary epidemic size at fixed $\rho = 0.5$ vs different values of the sporulation rate J_2 as obtained from simulation and predicted by the approximations.*

5.4.4 Comments

Whilst MF cannot describe any aspect of the anisotropic effects, PA seems to capture some qualitative features of the anisotropic NN model, although overall we find its predictions unsatisfactory; SA provides only a small quantitative improvement on PA and does not change the qualitative features which PA cannot capture (such as the shape of the phase diagrams). HPA provides the best descriptions of the model behaviour, both qualitatively, yielding the correct slope of the phase diagrams near the persistence threshold, and quantitatively, especially for some choices of parameters. In particular, in models including a small amount of background infection (which describe more realistic situations than models where the disease is transmitted only between neighbours), HPA gives extremely good predictions for the asymptotic behaviour. The description of the transient behaviour is more difficult and none of these approximations accurately predict the transient duration. HPA is, however, able to predict partially the slowing down of transients as anisotropy becomes stronger.

5.5 Conclusions

Epidemiological models typically assume that infectious spores are deposited isotropically around their source. In reality, a variety of factors affecting the

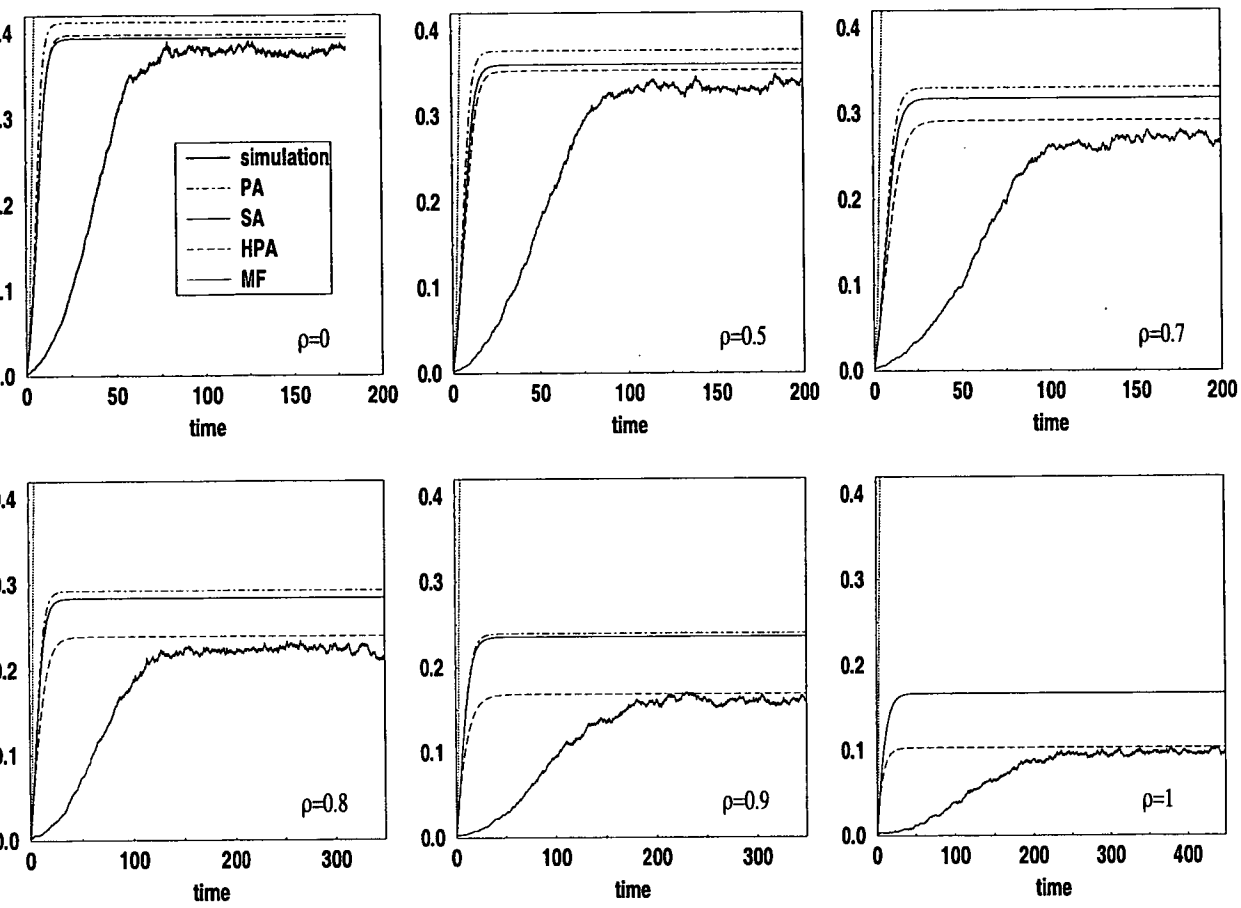


Figure 5.13: *Case I. Epidemic size vs time. Comparison of disease progress curves obtained from simulation and predicted by cluster approximations.* $J_2 = 2$, $J_1 = 0.01$, ρ *varying from* 0 *(isotropic model) to* 1 *(totally anisotropic model). Approximation curves from left to right (worst to best):* $\rho = 0$*: MF, PA, HPA, SA;* $\rho > 0$*: MF, PA, SA, HPA. Note that for* $\rho = 1$*, PA and SA coincide.*

spread of propagules might determine a preferential direction for the propagation of the disease. Prevailing winds, agricultural practices, water flows and field geometries are all possible causes for spatial anisotropy in the development of an epidemic. In this chapter we have formulated a generalisation of the NN model which allows for anisotropic spread of disease. We studied the behaviour of the model through simulation and analytical descriptions. Cluster approximations were tested for their ability to describe the effects of anisotropy.

In the anisotropic NN model, disease spreads from each infected to its four NNs with different strengths depending on the direction. We concentrated on the particular case in which there are only two different contact rates, along the vertical and the horizontal axes. The model can be thought of as representing

rectangular rather than square plots, where distances between rows and columns are different.

Despite the simplicity of the model, changes in behaviour resulting from the asymmetry in spread were clearly demonstrated using stochastic simulation. Two types of behaviour, corresponding to different values of the sporulation rate J_2, were examined. When J_2 is above the 1D persistence threshold, epidemics have large asymptotic levels and the effects of anisotropy on the stationary behaviour are very small. The epidemic size at equilibrium is almost constant, and only when the system becomes completely anisotropic (i.e. when disease can only be transmitted within each row) does the epidemic curve exhibit a sharp discontinuous transition. In this limit, the disease level depends on the initial number and spatial distribution of infecteds on the lattice. The duration of transients increases with the anisotropy. When J_2 is between the 1D and 2D persistence thresholds, the epidemic size gradually decreases as the degree of anisotropy increases. However, the system never becomes fully anisotropic as there is a threshold value of the anisotropy parameter above which the epidemic dies out. This case is more interesting from a biological point of view because it deals with a more realistic range of values for J_2.

Ordinary differential equations describing the dynamics of the anisotropic NN model were derived in an analogous way to the isotropic case. However, in this case the densities of differently oriented clusters are different and need to be considered as separate variables, therefore there are more equations. The closure approximations introduced in the previous chapter were applied to the model and compared to simulation.

MF is by construction non–spatial and isotropic and thus performs very poorly. PA and SA capture some qualitative features of the anisotropic behaviour but overall perform unsatisfactorily; in particular, SA does not provide a significant improvement on PA. On the other hand, HPA captures qualitative and quantitative effects of anisotropy far better than any other approximation.

All the approximations systematically underestimate the transient duration, in both the isotropic and anisotropic models. In particular, the duration of transients increases as the anisotropic effects increase, making transmission of the disease between rows more difficult: the approximations are much less accurate in predicting the effects of anisotropy on transients than on stationary disease level. The transient behaviour depends not only on the initial number of infecteds but also on their spatial distribution on the lattice. Since this information is not contained in the dynamic equations, the approximations are not expected to capture this aspect of the behaviour.

This chapter represents only a preliminary study of the effects of spatial anisotropy on disease spread. A deeper understanding could be gained from a more thorough simulation of the parameter space. It would be interesting to examine how the anisotropy affects the development of spatial correlations and how correlations along the horizontal and the vertical axes differ from each other. However, the model considered here is not the most general anisotropic NN model which could be formulated. The study could be extended to a model with four different deposition rates, which would be suitable to describe a larger range of spatial anisotropies and would have potentially interesting practical applications. Further down the line, it would be interesting to consider anisotropic extensions of more general dispersal models, such as those studied in chapters 3 and 6, where interactions span over all distances.

Chapter 6

Analytical methods for spatial models with general interactions

6.1 Introduction

In chapter 2 we derived a system of ODEs which describes the dynamics of a spatio–temporal model for the spread of plant epidemics. In this chapter we present closure approximations for closing and solving the system of equations. Their solutions are tested against the results of the simulation study in chapter 3.

The difficulty in obtaining analytical predictions for the behaviour of spatial stochastic models lies in the fact that such models are described by an infinite hierarchy of equations which is mathematically intractable. The usual solution is to investigate the behaviour using stochastic simulation. However, the simulation of models in which interactions are not restricted to local contacts is computationally very intensive and the inherent stochasticity of the model requires a large number of realisations in order to estimate the expected behaviour. For these reasons it is preferable to develop analytical descriptions offering deterministic predictions for the behaviour of the model.

Approaches based on cluster approximations have long been used in statistical physics (Bethe, 1935; ben-Avraham & Köhler, 1992), and more recently in population dynamics (Matsuda *et al.*, 1992; Levin & Durrett, 1997; Filipe & Gibson, 1998, 2000), for predicting the behaviour of lattice–based models characterised by nearest neighbour (NN) interactions. In chapters 4 and 5 we introduced a succession of orders of cluster approximations for such models in one and two spatial dimensions and tested their ability to capture the dynamics of isotropic and anisotropic NN epidemic models.

On the other hand, the rather simplistic, non–spatial Mean Field approximation (MF) has been widely used to study models with general interactions not

restricted to small neighbourhoods. However, the assumption that individuals are independent of each other leads to extremely poor predictions whenever significant spatial correlations develop in the distribution of the disease. Recently, Bolker & Pacala (1997, 1999) proposed an analytical method based on a spatial moment–closure (MC) approximation for predicting the expected evolution of population size. Such an approach is a potentially powerful tool for analysing the behaviour of stochastic spatial models. However, we found that its accuracy is restricted to situations in which interactions are weak and spatial correlations are small.

We propose a different approach built on individual–based dynamic equations and a range of closure approximations for predicting the behaviour of spatial models in which the individuals interact according to a generic function of their distance. The approximations, which include MC and more robust closure schemes capable of dealing with the development of spatial correlations, are applied to the general stochastic epidemic model introduced in chapter 2 and tested against some of the simulation results presented in chapter 3. In particular, we identify an approximation (the modified Kirkwood approximation, KA2) which provides very good predictions over most of the parameter space. This approach is both simple and robust and, in principle, can be applied to a broad range of models characterised by general interactions.

Part of the results presented in this chapter are reported in Filipe & Maule (2000 b, c).

6.2 Closure approximations

In chapter 2 we showed that the dynamics of the stochastic epidemic model considered are described by an infinite hierarchy of ODEs with no closed subsystems. Equation (2.21) for the time evolution of P_I involves P_{II}, equation (2.22) for P_{II} involves P_{III}, and so on. In order to solve the system (2.21)–(2.22) we need to close it. We do so by assuming approximate relations between higher and lower–order densities.

The usefulness of closure approximations goes beyond the mere purpose of obtaining deterministic predictions for the behaviour of the specific epidemic model considered here. Spatial models characterised by interactions between individuals invariably lead to nonlinear systems of equations which are open and therefore not solvable exactly. The closure schemes presented in this chapter can in principle be used for any spatial model characterised by general interactions.

The moment–closure approximation (MC) has recently been used to close

systems of ODEs describing the dynamics of various biological models. Bolker & Pacala (1997, 1999) applied this technique to a model for the growth and competition of plants in continuous and homogeneous space in which interactions decay exponentially with distance. Bolker (1999) considered a model describing the effects of dynamically–generated spatial heterogeneity on the development of plant epidemics. In that model the dispersal function was the modified Bessel function of order 0. Pascual & Levin (1999) investigated the consequences of density–dependent disturbance on a benthic population model in which interactions are represented by step functions. Those models are mathematically similar to the plant epidemic model considered here since they all include linear birth or death processes like primary infection and recovery, and nonlinear processes such as secondary infection.

Using the MC approximation, we found that the dynamic equations yield a solution only if spatial correlations in the population are weak and the behaviour of the model differs little from the corresponding non–spatial model. Hence, we investigated more robust approximations capable of dealing with strongly–correlated systems and applicable to spatial models under more general conditions.

In the rest of this section we review existing closure schemes, including MC, and propose alternatives. These are then applied to the system of dynamic equations (2.21)–(2.22).

Equation (2.22) represents a set of equations for $P_{II}(r,t)$, one for each possible distance r between individuals on a lattice. The number of equations is finite if the lattice is finite or if interactions are truncated.

We should recall that these equations describe a spatially–stationary and isotropic system with no boundaries in which all sites are statistically equivalent. In order to lighten the notation we shall omit the dependence on time and indicate distances in a symbolic way, so that xy represents the distance r_{xy} between sites $\mathbf{x}$ and $\mathbf{y}$, and xyz represents the relative distances r_{xy}, r_{xz}, r_{yz} between three sites $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$. Hence,

$$P_{II}(xy) = P_{II}(r_{xy}, t) \tag{6.1}$$

and

$$P_{III}(xyz) = P_{III}(r_{xy}, r_{xz}, r_{yz}, t). \tag{6.2}$$

We also denote the correlation function between two sites at distance r_{xy} by

$$C_{xy} = C(r_{xy}, t) = [P_{II}(xy) - P_I^2]/\sigma^2, \tag{6.3}$$

where $\sigma^2 = P_I - P_I^2$. Analogously to (2.17), we define the correlation function between three sites:

$$C_{\mathbf{xyz}} = \frac{\langle [S_{\mathbf{x}}(t) - \langle S_{\mathbf{x}}(t)\rangle]\ [S_{\mathbf{y}}(t) - \langle S_{\mathbf{y}}(t)\rangle]\ [S_{\mathbf{z}}(t) - \langle S_{\mathbf{z}}(t)\rangle]\rangle}{\sigma[S_{\mathbf{x}}(t)]\ \sigma[S_{\mathbf{y}}(t)]\ \sigma[S_{\mathbf{z}}(t)]}, \tag{6.4}$$

which, for a spatially–stationary system, becomes

$$C_{xyz} = C(r_{xy}, r_{xz}, r_{yz}, t) = \frac{P_{III}(xyz) - P_I\ [P_{II}(xy) + P_{II}(xz) + P_{II}(yz)] + 2P_I^3}{\sigma^2\ (1 - 2P_I)}, \tag{6.5}$$

where $P_{III}(xyz) = \langle S_{\mathbf{x}} S_{\mathbf{y}} S_{\mathbf{z}} \rangle$ and we used definitions (2.18) and (2.19).

6.2.1 Moment–closure approximation (MC)

From (6.5) we have

$$P_{III}(xyz) = \sigma^2\ (1 - 2P_I)\ C_{xyz} + P_I\ [P_{II}(xy) + P_{II}(xz) + P_{II}(yz)] - 2P_I^3. \tag{6.6}$$

Within the moment–closure approximation (MC) it is assumed that correlations between three sites are zero, i.e. $C_{xyz|MC} = 0$, hence

$$P_{III}(xyz) \approx P_{III}(xyz)_{|MC} = P_I\ [P_{II}(xy) + P_{II}(xz) + P_{II}(yz)] - 2P_I^3. \tag{6.7}$$

Substitution of (6.7) into (2.21)–(2.22) leads to an approximate closed system of equations.

The results presented by Bolker & Pacala (1997, 1999) were restricted to parameter regions where the rate of interactions between individuals is rather weak in comparison to the rate of linear processes. On the other hand, for the SI model in which there are no linear processes, Bolker (1999) considered epidemic models in which the initial fraction of randomly distributed infectives is large. Both features lead to a significant degree of spatial randomness and force correlations to be small. The MC system of equations was tested and found to yield no solution unless either the primary infection rate J_1 or the initial number of infecteds is sufficiently large.

The reason for the MC failure lies in the systematic underestimation of $P_{III}(xyz)$ in the equation for $dP_{II}(xy)/dt$ (2.22). Since $P_{III}(xyz)$ appears in (2.22) with a negative sign, by underestimating $P_{III}(xyz)$, MC overestimates $P_{II}(xy)$. An overestimation of $P_{II}(xy)$, which appears in the equation for dP_I/dt (2.21) with a negative sign, leads to an underestimation of P_I. In parameter regions where correlations are important, we verified that MC predicts a very fast increase in correlations in the first few steps of the integration, which makes P_I decrease without control. However, in parameter regions where MC does yield a solution, this solution is a good approximation.

6.2.2 Kirkwood approximation (KA)

The Kirkwood approximation (KA) was proposed by Kirkwood (1935) in the context of microscopic theories of liquids (Hansen & McDonald, 1979). KA assumes that

$$P_{III}(xyz) \approx P_{III}(xyz)_{|KA} = \frac{P_{II}(xy)P_{II}(xz)P_{II}(yz)}{P_I^3}, \tag{6.8}$$

which can be thought of as an extension of the usual pairwise approximation (PA, see e.g. Filipe & Gibson, 1998) to a situation in which the distances between the sites in the triplet are arbitrary. KA is a much more robust approximation than MC because it yields sensible solutions over the whole parameter space. However, we found that KA has undesirable features:

1. it systematically overestimates $P_{III}(xyz)$; consequently, following conversely the argument explained in the previous section, it overestimates P_I, especially in the case of MR and SR dispersals;

2. in the absence of primary infection the predicted phase diagram does not exhibit a clear persistence threshold.

6.2.3 Mixed approximations (KMC, KMix)

The two approximations considered so far lead to opposite errors in the estimation of P_I: MC underestimates the fraction of infecteds whilst KA overestimates it. This suggests mixing the two approximations and using a hybrid closure scheme. We did so in two different ways which lead to two different approximations.

The KMC approximation is defined by

$$P_{III}(xyz) \approx P_{III}(xyz)_{|KMC} = A\ P_{III}(xyz)_{|KA} + (1 - A)\ P_{III}(xyz)_{|MC}, \tag{6.9}$$

where $0 \leq A \leq 1$. We found that KMC with $A = 0.5$ always yields a better solution than KA. Using (6.5) and (6.7), we can in fact write:

$$C_{xyz|KMC} = \frac{A}{\sigma^2(1 - 2P_I)} \left[P_{III}(xyz)_{|KA} - P_{III}(xyz)_{|MC}\right] = A\ C_{xyz|KA}, \tag{6.10}$$

which shows that, for $0 \leq A < 1$, KMC predicts a smaller magnitude for the three-site correlation C_{xyz} than KA. From (6.6) it follows that KMC leads to smaller predictions for $P_{III}(xyz)$ and hence, using the same argument as above (see section 6.2.1), to smaller predictions for P_I.

From (6.6), (6.7), (6.8) and (6.3), we can also write

$$C_{xyz|KA} = \frac{K_1(xyz) + K_2(xyz)}{\sigma^2(1 - 2P_I)}, \tag{6.11}$$

where

$$K_1(xyz) = (1 - P_I)^3 \; C_{xy}C_{xz}C_{yz} \tag{6.12}$$

and

$$K_2(xyz) = P_I(1 - P_I)^2 \; [C_{xy}C_{xz} + C_{xy}C_{yz} + C_{xz}C_{yz}]. \tag{6.13}$$

The two terms $K_1(xyz)$ and $K_2(xyz)$ have different effects on the transient and stationary behaviour, suggesting another mixing approximation, which we named KMix:

$$P_{III}(xyz) \approx P_{III}(xyz)_{|KMix} = \sigma^2(1 - 2P_I)C_{xyz|KMix} + P_{III}(xyz)_{|MC}, \tag{6.14}$$

where

$$C_{xyz|KMix} = BK_1(xyz) + (1 - B)K_2(xyz). \tag{6.15}$$

We found that KMix with $B = 0.8$ yields an optimal mixing approximation which always improves the KMC and KA solutions.

However, like KA, neither of these two mixing approximations exhibit a clear persistence threshold. Moreover, the arbitrariness in the choice of the mixing parameters A and B may be seen as a weakness of these closure schemes.

6.2.4 Modified Kirkwood approximation (KA2)

Equations (2.21) and (2.22) can be rewritten as

$$\frac{dP_I}{dt} = J_1 - (J_1 + R) \; P_I + J_2 \sum_{\mathbf{y}(\neq \mathbf{x})} P_{SI}(xy) \; f(r_{xy}) \tag{6.16}$$

and

$$\begin{aligned} \frac{1}{2}\frac{dP_{II}(xy)}{dt} = & \; [J_1 + J_2 \; f(r_{xy})] \; P_I - [J_1 + J_2 \; f(r_{xy}) + R] \; P_{II}(xy) \\ & + \qquad J_2 \sum_{\mathbf{z}(\neq \mathbf{x},\mathbf{y})} P_{SII}(xyz) \; f(r_{xz}), \end{aligned} \tag{6.17}$$

where $P_{SI}(xy) = P_I - P_{II}(xy)$ is the probability of finding a susceptible–infected (SI) pair at distance r_{xy} apart, and $P_{SII}(xyz) = P_{II}(yz) - P_{III}(xyz)$ is the probability of finding a triplet of sites (susceptible–infected–infected, SII) at relative

distances r_{xy}, r_{xz} and r_{yz}. Written in this form, the dynamic equations (6.16) and (6.17) suggest a modification of KA.

We noted above that KA can be thought of as an extension of PA. We also recall that, for a triplet of adjacent sites where the central site is susceptible (denoted by ISI), PA assumes that $P_{ISI|PA} = P_{SI}^2/P_S$ (Filipe & Gibson, 1998). A natural way of generalising this approximation to a situation in which the three sites may be positioned at any distance from each other is to assume that:

$$P_{SII}(xyz) \approx P_{SII}(xyz)_{|KA2} = \frac{P_{SI}(xy)P_{II}(yz)P_{SI}(xz)}{P_S P_I^2}, \tag{6.18}$$

where $P_S = 1 - P_I$. We call this approximation the modified Kirkwood approximation (KA2).

We found that KA2 substantially outperforms all the above approximations and most of the existing approximations for NN models (Filipe & Gibson, 2000) over the whole parameter space. In addition, KA2 has none of the disadvantages of the other approximations discussed here; in particular, in the absence of primary infection, it exhibits a clear persistence threshold.

Some properties of this closure scheme can be derived analytically.

1. Both KA and PA tend to overestimate the fraction of infecteds. We now show that the KA2 prediction for P_I is smaller than the KA and PA predictions (Filipe & Maule, 2000 b). From (6.8) and (6.18) we have

$$P_{III}(xyz)_{|KA2} = P_{III}(xyz)_{|KA} - (\sigma/P_I)^2\, C_{xy}C_{xz}P_{II}(xy), \tag{6.19}$$

hence $P_{III}(xyz)_{|KA2} < P_{III}(xyz)_{|KA}$. The same argument explained in section 6.2.1 tells us that a smaller prediction for P_{III} implies a smaller prediction for P_I (see equation (2.21)). In the special case of the NN model, PA assumes that $P_{ISI}(zxy)_{|PA} = P_{IS}(zx)P_{SI}(xy)/P_S$, provided that $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are three adjacent sites. Hence,

$$P_{ISI}(zxy)_{|KA2} = P_{ISI}(zxy)_{|PA}(P_{II}(yz)/P_I^2). \tag{6.20}$$

Since $P_{II}(yz) \geq P_I^2$, then $P_{ISI}(zxy)_{|KA2} \geq P_{ISI}(zxy)_{|PA}$, and a larger prediction for P_{ISI} implies a smaller prediction for P_I. Therefore KA2 should yield an improvement on KA in the general case and on PA for the NN model.

2. It can be shown that $P_{III}(xyz)_{|KA2} > 0$, at least in the extreme cases of maximum and minimum correlations (Filipe & Maule, 2000 b). When correlations are zero we have, from (6.19), $P_{III}(xyz)_{|KA2} = P_{III}(xyz)_{|KA} > 0$. When correlations are maximum, we have $P_{II} = P_I$; substituting the latter expression in (6.8) and (6.19) we obtain $P_{III}(xyz)_{|KA2} = P_I > 0$.

6.3 Algorithm implementation

In order to solve the system of dynamic equations (2.21)–(2.22) we use a closure approximation to truncate the system at a particular order (in this case at the second order) and implement an algorithm to obtain a numerical solution.

We considered two different approaches, namely a *lattice* and a *continuous* approach.

In the lattice approach we think of equations (2.21) and (2.22) as representing the processes at the level of individuals exactly as in the simulation. In particular, (2.22) comprises a set of equations where r_{xy} equals any of the possible distances between lattice sites. This approach is different from the one adopted by Bolker & Pacala (1997, 1999), who derived and solved an integro–differential system of equations in continuous space. In the lattice implementation, the dynamic equations are simple, transparent and easy to solve, and preserve the spatial discreteness of individuals. Moreover, an approximation which works well on a lattice is likely to work well using any other distributions of discrete hosts. A drawback to this approach, however, is the number of equations of the system, which becomes very large when we consider large lattices. For example, the number of distances between 1 and 50 on a square lattice is 761; thus (2.22) represents a set of 761 equations.

We also consider an approach in which the densities P_I and $P_{II}(r)$ are not specified at lattice sites but over a mosaic of quadrats defined using polar co-ordinates, i.e. the densities are assumed to be constant over each quadrat. The quadrat structure bears no relation to the lattice; it is used for computational convenience and to eliminate the explicit consideration of a lattice. The approach is analogous to spatial discretisation schemes used to solve numerically some continuous–media problems; for this reason we call this quadrat approach the "continuous approach". We consider this approach for two reasons: (1) to try and reduce the number of equations in the system (2.21)–(2.22), and (2) to extend the applicability of the current approach to models in which hosts are not distributed on a lattice; this includes both the cases of discrete hosts in continuous space and a continuous mass–density of hosts (as assumed in the integro–differential approach of Bolker & Pacala (1997, 1999)). In order to reduce the number of equations in the system we exploit the fact that, above moderate distances between sites on the lattice (e.g. $r > 5$), there are many pairs of sites at very similar distances; since $P_{II}(r)$ is expected to vary slowly with r, the quadrat approach should provide a good approximation at least for these distances. We expect this approach to be less accurate for short–range interactions: in this

case, most infections occur over short distances and this makes behaviour more dependent on the lattice structure, which is ignored in this approach.

6.4 Lattice approach

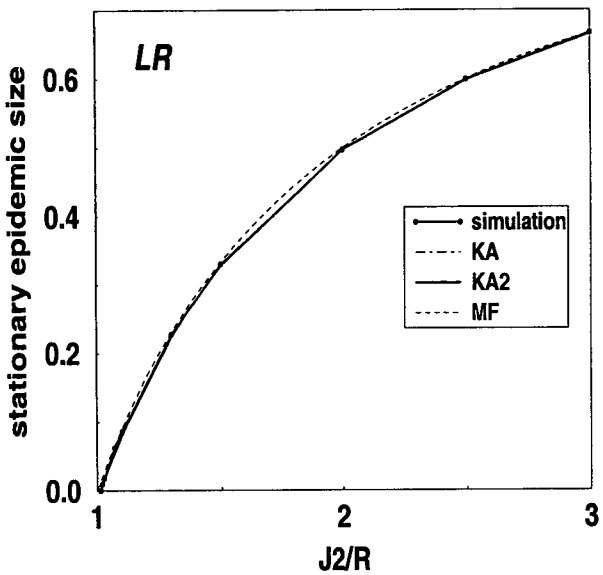


Figure 6.1: *LR dispersal. Phase diagram obtained from simulation and predicted by MF, KA and KA2 (note that simulation and approximation curves are indistinguishable).*

Within the lattice approach we let r_{xy} and r_{xz} in (2.22) be lattice distances between 1 and $L/2$, with L the size of the simulated system. When $r_{yz} > L/2$, we apply the MF approximation to the probability $P_{II}(yz)$ (i.e. $P_{II}(yz) \approx P_I^2$ if $r_{yz} > L/2$); since L is large, this has a negligible effect. We consider a lattice of size $L = 100$; we have verified that using $L = 200$ produces no visible changes. For this L, the system of ODEs (2.21)–(2.22) consists of 762 equations. We solved the equations using a basic numerical routine (D02CJF) from the Numerical Algorithms Group (1993) which solves the initial value problem for a first order system of equations using a variable–order variable–step Adams method (see, for example, Hall & Watt, 1976). CPU performance on a SUN Ultra 2170 workstation varies between seconds and about one hour, depending on parameter values. Despite the large number of equations, convergence to a solution is very rapid because many of the lattice distances are very similar. On the other hand, simulation of single realisations varies between 1 and 48 hours and estimation of expectations might require hundreds of realisations (see Filipe & Maule, 2000 a). The computational benefits of the analytical approach are thus evident.

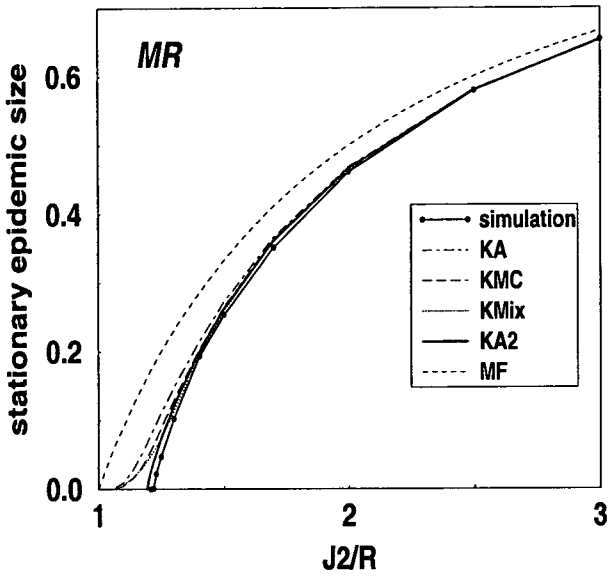


Figure 6.2: *MR dispersal. Phase diagram obtained from simulation and predicted by the closure approximations. Approximation curves from left to right (worst to best) near the threshold: MF, KA, KMC, KMix, KA2. Note that for some values of J_2/R (away from the threshold) KMC and KMix perform better than KA2.*

The approximation predictions are tested against simulation of the epidemic model in cases of LR, MR and SR dispersal and different sporulation rates J_2. Since closure approximations generally underestimate correlations, the most demanding scenarios for testing them correspond to parameter regions where strong correlations develop. Maximum correlations occur in the absence of primary infection ($J_1 = 0$). We shall focus first on this situation, which represents a lower bound on the performance of the approximations. Since in real systems it is likely that the spread of disease is a combination of primary and secondary infection, later we shall consider the presence of a small amount of background infection. The two versions of the model (with and without recovery of infecteds) are analysed below.

6.4.1 Results. SIS model

6.4.1.1 Stationary behaviour

Figs 6.1, 6.2 and 6.3 show the long–term epidemic size against infectiousness J_2/R, with $J_1 = 0$, in LR, MR and SR dispersal cases, respectively. In the SR case (Fig. 6.3), curves corresponding to existing cluster approximations for models with SR interactions (PA, SA and HPA, see chapter 4) are also shown. Unsurprisingly, in the LR case the approximations (including MF) and simulation are indistinguishable, whilst in the SR case predictions are less effective. In the MR

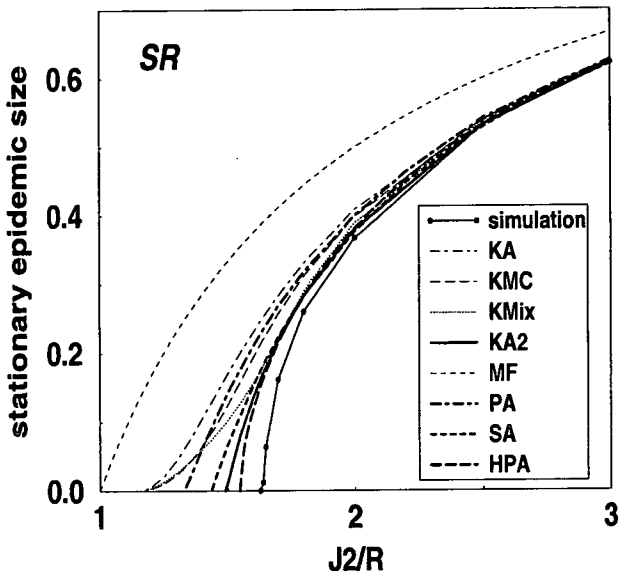


Figure 6.3: *SR dispersal. Phase diagram obtained from simulation and predicted by the closure approximations. Approximation curves from left to right (worst to best) near the threshold: MF, KA, KMix, KMC, PA, SA, KA2, HPA. Note that KMC and KMix perform particularly poorly near the threshold but outperform PA and SA for large values of* J_2/R.

case, which might be the more realistic, predictions are very good. In all dispersal cases KA2 performs best whilst KA performs worst. The KMix approximation does particularly well in some regions but, like KMC and KA, performs poorly at low epidemic levels. In the SR case (Fig. 6.3) KA2 overperforms PA and SA (Filipe & Gibson, 2000) and is only inferior to the more elaborate HPA (Filipe, 1999 a, b). No solutions were obtained with the MC approximation for $J_1 = 0$.

In chapter 2 we showed that, for infinite systems with LR interactions ($a \leq 2$), MF represents the exact solution to the model. Fig. 6.1 shows that MF does in fact accurately predict the stationary epidemic size. The advantage of using spatial approximations rather than MF in the LR case lies in the fact that they can also predict spatial correlations, which are completely neglected by MF (Figs 6.4a and b).

Figs 6.4, 6.5, 6.6 show simulated and predicted correlations for LR, MR and SR and for two levels of infection ($10\%, 25\%$). As the epidemic size or the range of interaction decrease, correlations become larger and the efficacy of the approximations diminishes. Correlations are underestimated both in magnitude and range for SR and MR, whilst in the LR case correlations are overestimated (note that all approximation curves overlap in Fig. 6.4). Underestimation of spatial correlations results in the observed overestimation of the stationary epidemic

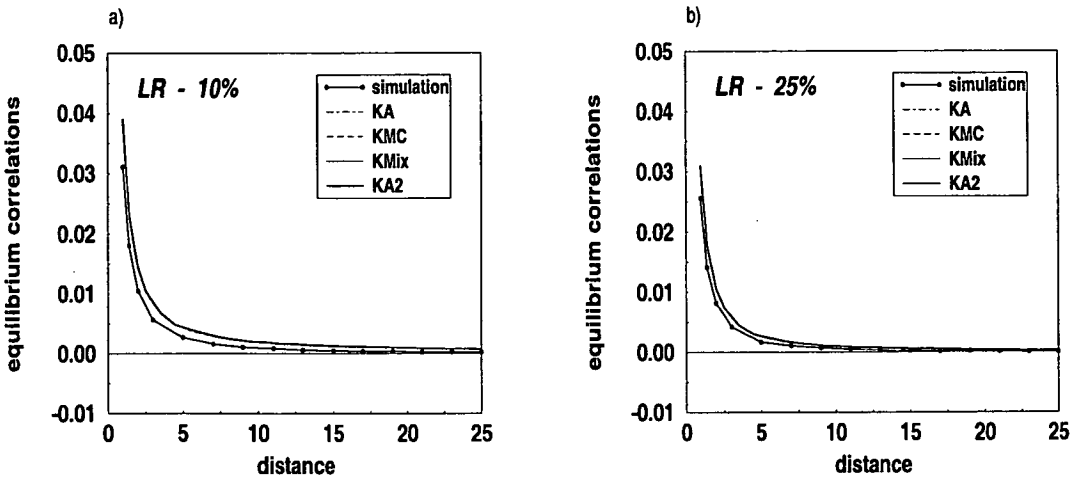


Figure 6.4: *Equilibrium correlations obtained from simulation and predicted by the closure approximations (note that all approximation curves are indistinguishable). LR, a)* $P_I(\infty) = 10\%$*; b)* $P_I(\infty) = 25\%$.

interaction	$P_I(\infty)_{sim}$	$C(1)_{sim}$	$C(1)_{KA}$	$C(1)_{KMC}$	$C(1)_{KMix}$	$C(1)_{KA2}$
LR	10%	0.0311	0.0383	0.0388	0.039	0.039
	25%	0.0256	0.0307	0.0310	0.0310	0.0310
MR	10%	0.1926	0.1381	0.1607	0.1777	0.1705
	25%	0.1470	0.1193	0.1303	0.1394	0.1354
SR	10%	0.3381	0.1798	0.2100	0.2457	0.2503
	25%	0.2704	0.1671	0.1884	0.2116	0.2192

Table 6.1: *Values of C(1) obtained from simulation and predicted by the closure approximations for LR, MR and SR interactions and for two stationary epidemic levels.*

size (see Figs 6.2, 6.3). Table 6.1 shows the values of the spatial correlation at lag 1, $C(1)$, corresponding to the cases illustrated in Figs 6.4, 6.5 and 6.6.

We noted above that in some parameter regions KMix performs very well and provides the best approximations for long–term epidemic size (see Figs 6.2 and 6.3). Fig. 6.5 shows very good agreement between simulated correlations and KMix predictions. However, some problems occur with this approximation at lower levels of infection. The spatial correlations predicted by KMix for small disease levels do not converge to 0 but to a small positive value. Fig. 6.7 shows an example for the MR case with $J_2 = 1.22$, corresponding to a stationary epidemic size of ~ 0.001 (the corresponding epidemic size predicted by KMix is 0.055). Thus, there are anomalies in the performance of KMix which make it a weaker approximation by comparison to the others.

The same problem occurs with the KMC approximation with values of the

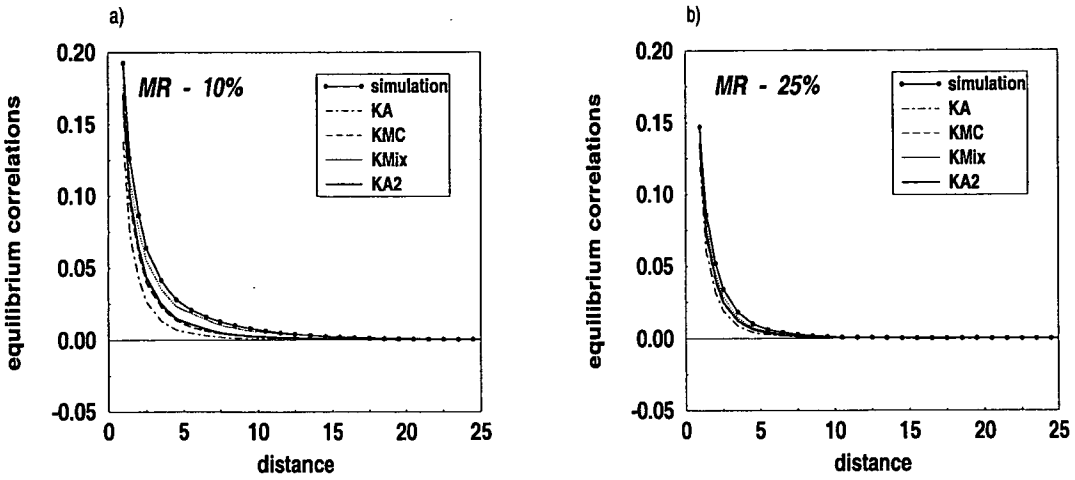


Figure 6.5: *Equilibrium correlations obtained from simulation and predicted by the closure approximations. MR, a) $P_I(\infty) = 10\%$; b) $P_I(\infty) = 25\%$. Approximation curves from left to right (worst to best): KA, KMC, KA2, KMix.*

mixing parameter $A < 0.5$. Lower values of A, in fact, provide better predictions for the long–term disease level, but in the proximity of the persistence threshold predict correlations which converge to a positive value.

6.4.1.2 Transient behaviour

Next we investigate the ability of the closure approximations to capture transient behaviour. Fig. 6.1 showed that all the approximations, including MF, give very good predictions for the long–term behaviour of the LR dispersal model. They also provide good predictions for the transient behaviour of the LR model, so we focus here on MR and SR dispersal cases. Figs 6.8 and 6.9 show simulations of epidemic transients for three stationary disease levels $(10\%, 25\%, 50\%)$, with $J_1 = 0$, and the corresponding predictions by MF, KA, KA2 and, for the SR case, PA and HPA. Epidemics are initiated with a small fraction of randomly distributed infecteds ($P_I = 0.2\%$), as in chapter 3.

Conclusions analogous to those above can be drawn: MF always offers very poor descriptions, whilst approaches with some spatial information, such as KA and PA (in the SR case), capture transient behaviour better, especially in parameter regions where correlations are not too strong. In the MR case (Fig. 6.8) predictions are very reasonable, especially those of KA2. As expected, predictions are less accurate in the SR case. However, even here KA2 outperforms PA and SA (not shown) and, in the range of parameters considered, is not inferior to HPA (note that in Fig. 6.9b the two curves are indistinguishable).

A common feature, however, is that all the approximations underestimate the duration of transients, especially when long–term values are low and transients

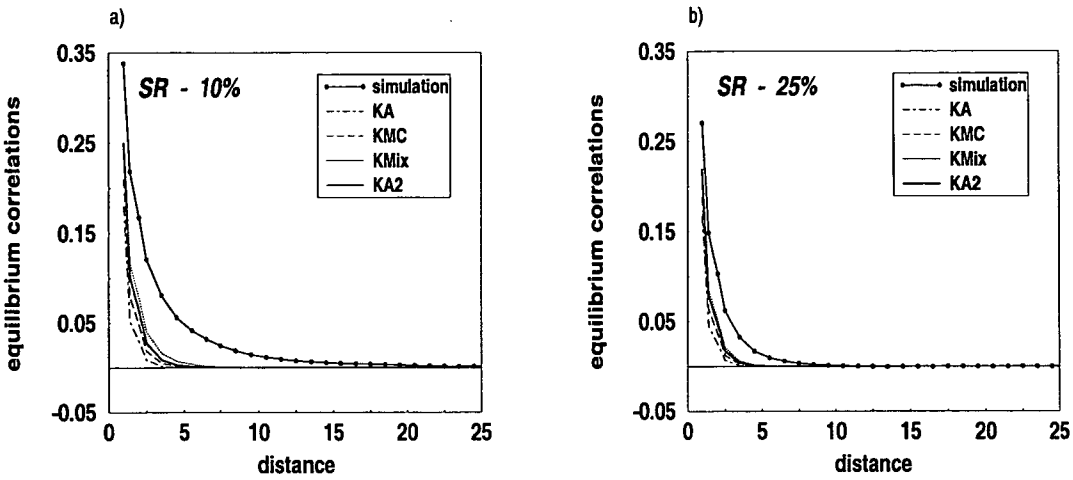


Figure 6.6: *Equilibrium correlations obtained from simulation and predicted by the closure approximations. SR, a) $P_I(\infty) = 10\%$; b) $P_I(\infty) = 25\%$. Approximation curves from left to right (worst to best): KA, KMC, KA2, KMix.*

are slow. As already noted in chapter 5 with regard to the performance of cluster approximations in the isotropic and anisotropic NN models, closure approximations are generally more successful in predicting stationary rather than transient behaviour (see sections 5.4.1 and 5.5). We believe that the reason for this difference is the absence of spatial stationarity in the initial states and therefore in the transients, whilst the ODEs used to describe the spatially–explicit model assume spatial stationarity at all times.

All the approximations are expected to give better predictions when the model includes background infection. In this case the heterogeneity in the distribution of the disease is reduced and correlations are smaller. From a biological point of view, primary infection makes the model more realistic and, from the point of view of testing different approximations, gives us the opportunity to apply and test the MC approximation. However, if the rate of primary infection rate is too large, this linear process dominates and the spatial component of the model behaviour is less important. In this case the use of spatial approximations is not essential since even MF may give satisfactory predictions with minimal computational effort. For example, in Fig. 6.10 MF, KA and MC are compared to a simulated transient in the case of MR dispersal with rates of infection $J_1 = 0.1, J_2 = 1.48$ (i.e. the linear process is roughly 15 times weaker than the interaction process). For these parameter values both KA and MC offer very good predictions, although they actually differ very little from the non–spatial MF curve.

For a given set of initial conditions (defined by $P_I(0) = 0.2\%$) and a given J_2, there is a threshold value of J_1 below which MC produces anomalous solutions before breaking down, as explained below. We focus on values of J_2 which would

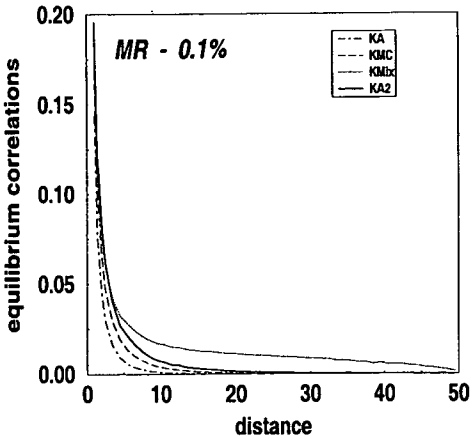


Figure 6.7: *Approximation predictions for stationary spatial correlations in the proximity of the critical threshold (MR). Note the anomalous behaviour of KMix by comparison to the other curves.*

generate epidemics of size 10% in equilibrium if there were no background infection ($J_2 = 1.3$ for MR and $J_2 = 1.688$ for SR). We find that the minimum values of J_1 for which MC yields a solution are $J_1 = 0.02$ and $J_1 = 0.05$ for MR and SR, respectively. Hence, the smaller the range of dispersal, the larger the rate of the primary infection needed to obtain a MC solution. Figs 6.11a and b show the corresponding transients: there is little difference between spatial approximations and simulation, whilst the non–spatial MF approximation significantly overestimates the epidemic size. In the MR case, MC slightly underestimates the epidemic size.

For values of J_1 below this threshold, MC first exhibits an inflection, reflecting a numerical instability, and then breaks down, as shown by the two bottom curves in Fig. 6.11a. For even lower values of J_1, (Fig. 6.11c), MC yields no solution, while some of the other approximations give good predictions for both the epidemic size and the transient duration. KA2 again proves to be the best closure scheme.

6.4.2 Results. SI model

We now analyse the ability of closure approximations to capture the behaviour of epidemic models which do not include the recovery of infecteds. Obviously in this case there is no stationary regime. In this section we restrict our attention to the case $J_1 = 0$. As before, epidemics are initiated with a small fraction of randomly distributed infecteds ($P_I(0) = 0.2\%$). Fig. 6.12 compares analytical predictions by MF and KA2 to simulation of disease progress curves in the cases of LR, MR and SR dispersal. No solutions were obtained with the MC approximation for

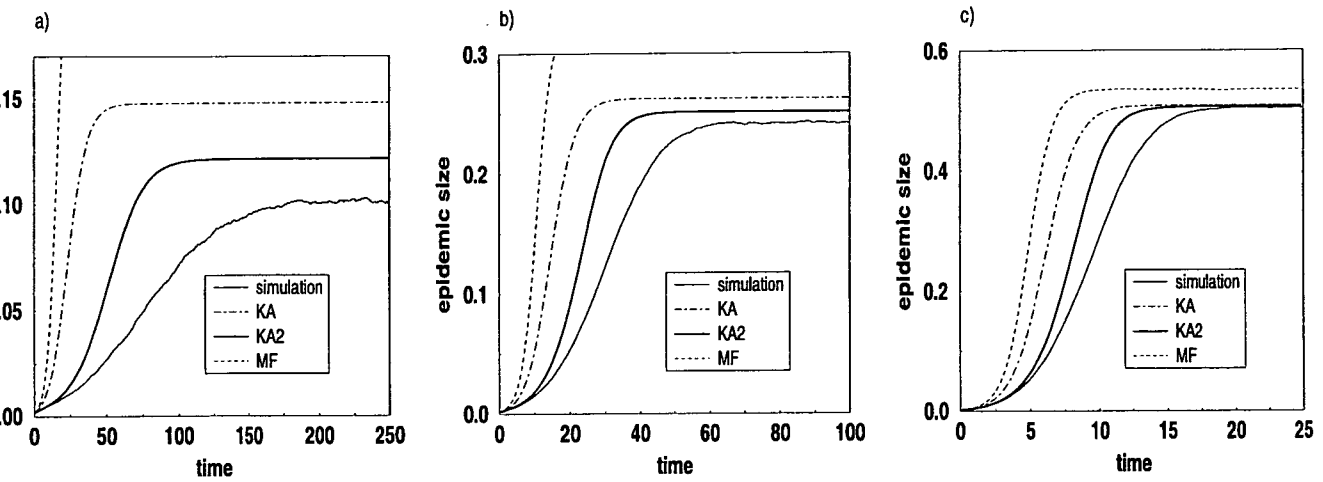


Figure 6.8: *Expected epidemic transients in the MR dispersal case. Long term epidemic size: a)* 10%, *b)* 25%, *c)* 50%.

this initial condition.

With LR dispersal, simulation and KA2 prediction are indistinguishable, whilst MF slightly overestimates the disease level (Fig. 6.12a). MF strongly underestimates the durations of MR and SR transients, whilst KA2 yields a good prediction in the MR case and significantly improves MF and PA in the SR case (Figs 6.12b and c). It should be noted that PA and HPA coincide for the SI model (see (4.62)).

In a recent paper, Bolker (1999) used the MC approximation to truncate the hierarchy of dynamic equations. That study was made in the context of SI and SIR models, and MC predictions were compared to simulation where discrete hosts were distributed at locations in continuous space. Bolker investigated both the cases of random and slightly clustered host distributions. We wish to compare Bolker's results (concerning the SI model and the non–clustered host distribution) to ours. To this end, in addition to the usual power–law contact distributions, we shall also consider a dispersal function given by the modified Bessel function of order 0, $K_0(\psi r)$, where the length scale $1/\psi$ is a measure of the range of the dispersal. Following Bolker (1999), we set $\psi = 0.5$, which corresponds to long–range dispersal, as we verified by comparing disease progress curves. The asymptotic behaviour of Bessel functions of this type is exponential ($K_0(\psi r) \underset{r\to\infty}{\simeq} \exp(-\psi r)/\sqrt{2\pi\psi r}$). Since our model is lattice based, the contact distribution is normalised over all possible depositions on the dispersal domain, as defined in chapter 2 (see (2.3) and (2.4)).

As already noted, MC yields solutions to the system of equations only when spatial correlations in the distribution of the disease are small. In the context

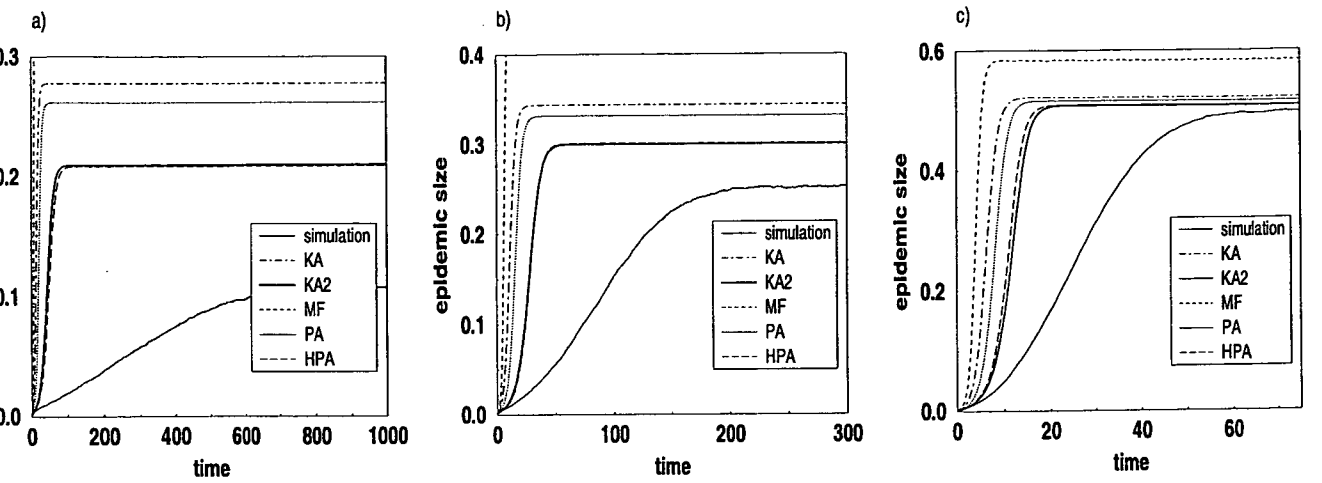


Figure 6.9: *Expected epidemic transients in the SR dispersal case. Long term epidemic size: a)* 10%, *b)* 25%, *c)* 50%. *Approximation curves from left to right (worst to best): MF, KA, PA, KA2, HPA. Note that KA2 and HPA curves are very close: they coincide for* $\mu = 25\%$, *HPA performs better than KA2 for* $\mu = 10\%$, *and worse for* $\mu = 50\%$.

of an SI model with no primary infection (as in Bolker, 1999), we found that one requirement for MC to provide a solution is to have a large initial density of randomly distributed infecteds. Such an initial condition has a persistent randomising effect on the subsequent spatial dynamics. The disease progress curves shown by Bolker (1999) (corresponding to Bessel dispersal functions) start with a fraction of 5% infected hosts (Figs 2a and 4a in Bolker, 1999). Large initial levels of random infection are in general unrealistic, have strong effects on the transient behaviour, and noticeably diminish stochastic variability. Given that this condition is mathematically necessary for the performance of the MC approximation, there are limitations in the range of situations to which MC can be applied. Fig. 6.13 shows how two simulated transient epidemics with SR dispersal differ when the initial conditions consist of a fraction of 0.2% (right–hand curve) and 10% (left–hand curve) randomly distributed infecteds. Since the other closure approximations presented in this chapter can be applied to any initial condition (Fig. 6.12), they offer valid alternatives to MC.

Fig. 6.14 compares analytical predictions and simulation for the minimum initial levels of infection, corresponding to LR, MR and SR respectively, for which MC yields a solution. It should be noted that the LR dispersal case is represented by a contact distribution based on the modified Bessel function with parameter $\psi = 0.5$, chosen in order to reproduce the conditions in Bolker's paper (1999) as much as possible. The initial fractions of infected are $P_I(0) = 10\%$ in the SR case and $P_I(0) = 5\%$ in the MR and LR cases (although in the LR case correlations are quite small and MC yields a solution for initial infections as low

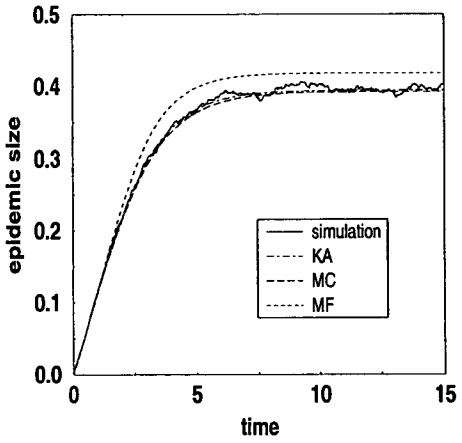


Figure 6.10: *Comparison between non-spatial (MF) and spatial (KA, MC) approximation predictions for an epidemic developing according to MR dispersal and large primary infection rate ($J_1 = 0.1, J2 = 1.48$). Note that KA and MC curves are very close.*

as 1% of infecteds). KA2 and MC perform quite well in all dispersal cases: these predictions are indistinguishable for LR and differ little for MR and SR, although MC systematically underestimates the disease level. MF underestimates transient durations slightly; nevertheless, with these initial conditions, even MF predictions differ little from simulation.

The lower curves in Fig. 6.14 show the same kind of instability observed in the case of the SIS model (Fig. 6.11a), which occurs just before MC breaks down. They correspond to the following initial conditions: $P_I(0) = 0.6\%$, 3% and 7.5% for LR (Bessel function), MR and SR dispersal, respectively.

We recall that Γ (see (2.31)) represents a convolution of the correlation and dispersal functions. Fig. 6.15 shows the time evolution of $\Gamma\sigma^2$ as predicted by KA2 and MC and obtained from simulation for LR, MR and SR dispersal cases. The initial fraction of infecteds is $P_I(0) = 5\%$, as in Bolker (1999). In Bolker (1999) this quantity was denoted by $\bar{c}_{II}$ and called *average covariance density* (see Fig. 2c in that paper). As expected, for LR dispersal spatial correlations, and hence Γ, are very small, and both KA2 and MC provide good predictions. In the MR case KA2 underestimates the covariance whilst MC overestimates it. In the SR case the covariance is maximum and MC does not yield any solution (the minimum value of $P_I(0)$ for which a MC solution was obtained is 10%).

Fig. 6.15a, which shows the evolution of $\bar{c}_{II}$ (covariance of infectives), should be compared to the plot of $\bar{c}_{SS}$ (covariance of susceptibles) in Fig. 2c of Bolker (1999). For our model, defined on a regular lattice, we have $\bar{c}_{II} = \bar{c}_{SS}$ and $\bar{c}_{SI} = -\bar{c}_{II}$, whilst with a general spatial distribution of hosts (Bolker, 1999) the relation between covariances is given by $\bar{c}_{II} + \bar{c}_{SS} = -2\bar{c}_{SI}$. This reduces to the

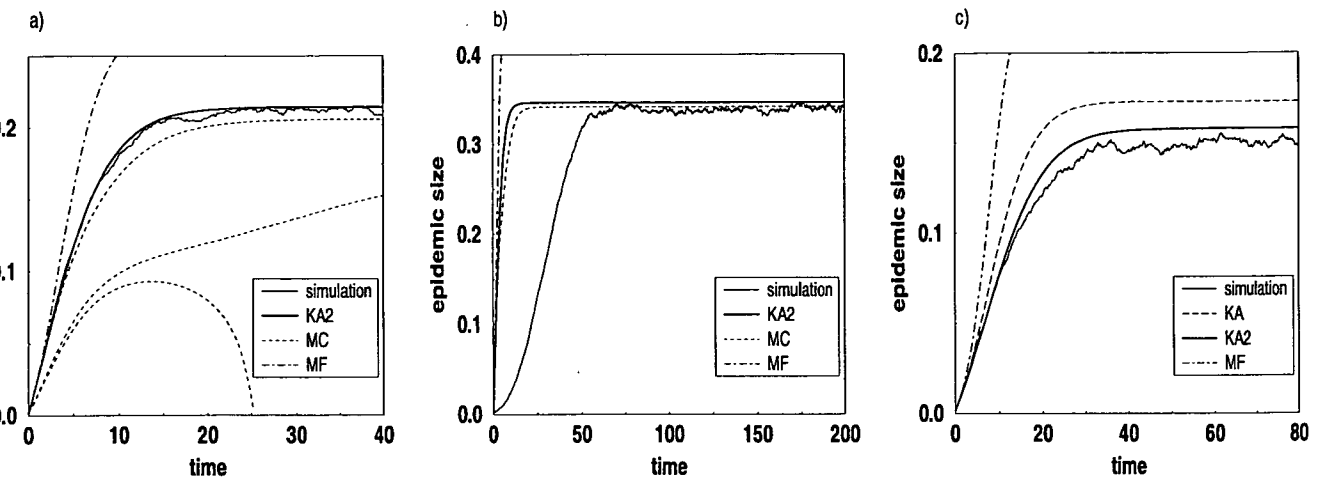


Figure 6.11: *Expected epidemic transients with a small rate of primary infection. Secondary infection rate J_2 corresponds to 10% long-term disease levels when $J_1 = 0$. a) MR, $J_1 = 0.02, J_2 = 1.3$. MF overestimates P_I, KA2 and simulation curves are very close, and MC slightly underestimates P_I. The two bottom curves correspond to values of J_1 for which MC breaks down: $J_1 = 0.011$ and 0.01; b) SR, $J_1 = 0.05$, $J_2 = 1.688$. Approximation curves from left to right (worst to best): MF, KA2, MC; c) MR, $J_1 = 0.005$, $J_2 = 1.3$. Approximation curves from left to right (worst to best): MF, KA, KA2. Note that MC yields no solution.*

former expressions for a lattice distribution of hosts (Filipe, 2000, unpublished).

6.5 Continuous approach

In this approach continuous space is discretised by considering a mosaic of quadrats within which the densities P_I and $P_{II}(r)$ are assumed to be constant. Fig. 6.16 shows an example of how a portion of space is subdivided into 18 quadrats.

The system of ODEs (2.21)–(2.22) applies to a population distributed on a lattice. In order to solve it within the continuous approach, the equations must be reformulated according to this new discretisation of space.

By using the radial distribution of sites $G(r)$ (see section 2.3), equation (2.21) can be rewritten as

$$\frac{dP_I}{dt} = J_1 - (J_1 - J_2 + R)\ P_I - J_2 \sum_{r \in \mathcal{L}_{1,L/2}} P_{II}(r) G(r) f(r), \qquad (6.21)$$

where $\mathcal{L}_{1,L/2} = \{1, \sqrt{2}, 2, \sqrt{5}, \ldots, L/2\}$ is the set of all lattice distances between 1 and $L/2$, with L the size of the simulated system. We now define another set of distances, $\mathcal{L}'_{1,L/2} = \{1, 1+\Delta, 1+2\Delta, \ldots, L/2\}$, where Δ is a small constant

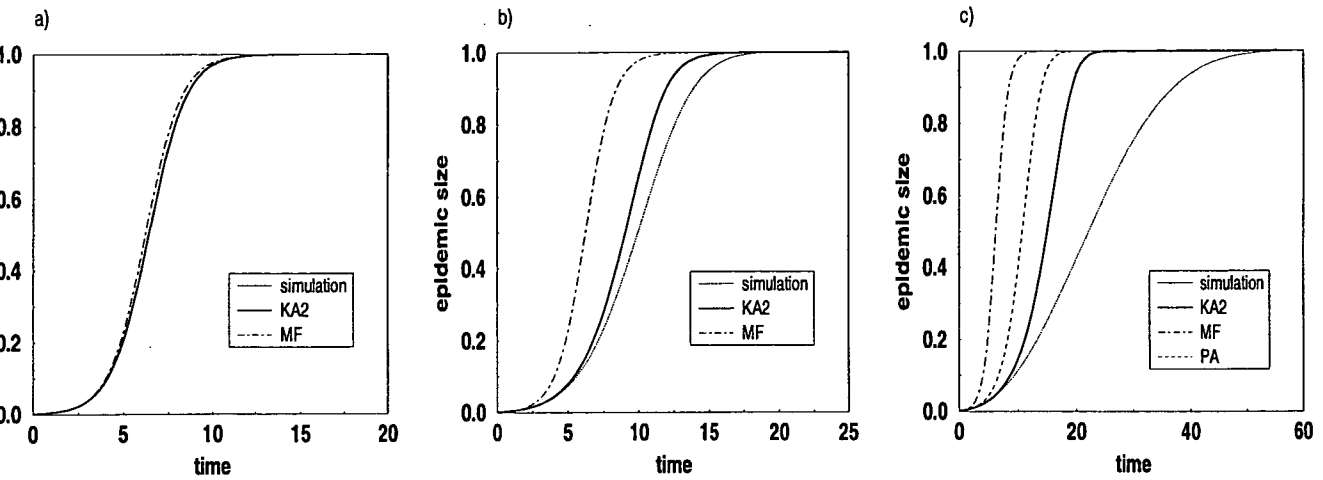


Figure 6.12: *Expected disease progress curves with initial condition $P_I(0) = 0.2\%$. a) LR. Simulation and KA2 curves are indistinguishable, MF overestimates P_I; b) MR. Approximation curves from left to right (worst to best): MF, KA2; c) SR. Approximation curves from left to right (worst to best): MF, PA, KA2. Note that MC yields no solution in all the dispersal cases.*

interval (see Fig. 6.16). Equation (6.21) can then be rewritten as

$$\frac{dP_I}{dt} = J_1 - (J_1 - J_2 + R)\ P_I - J_2 \sum_{s \in \mathcal{L}'_{1,L/2}} \sum_{r \in \mathcal{L}_{s,s+\Delta}} P_{II}(r) G(r) f(r), \qquad (6.22)$$

where $\mathcal{L}_{s,s+\Delta}$ is the set of lattice distances in the interval $[s, s+\Delta[$. Since $P_{II}(r)$ is expected to be smooth and to vary little in the interval $r \in [s, s+\Delta[$, we assume that $P_{II}(r)$ is constant over this interval. This approximation depends on the length Δ and is expected to be less accurate for small r and SR interactions. Ignoring the variation of $P_{II}(r)f(r)$ in the interval $[s, s+\Delta[$, we can extract this factor from the second sum in (6.22) and write:

$$\frac{dP_I}{dt} = J_1 - (J_1 - J_2 + R)\ P_I - J_2 \sum_{s \in \mathcal{L}'_{1,L/2}} P_{II}(s) f(s) H(s, \Delta), \qquad (6.23)$$

where $H(s, \Delta) = \sum_{r \in \mathcal{L}_{s,s+\Delta}} G(r)$ is the histogram of $G(r)$ with window of width Δ (see (2.7)). The set of distances $\mathcal{L}'_{1,L/2}$ in (6.23) is much smaller than $\mathcal{L}_{1,L/2}$ in (6.21) because the density of points on a lattice increases with r.

Analogously, we wish to rewrite the equation for the evolution of $P_{II}(r)$. We use $\Theta_r = \{\theta_1, \theta_2, \ldots, \theta_{G(r)}\}$ to denote the set of angular coordinates of the lattice sites at distance r from a given site. For example, in Fig. 6.17 the dark dots (•) indicate the locations of lattice sites at distance $r = 5$ from a given site (located at the centre of the circle). θ_1 and θ_2 are the angular coordinates of two of these sites.

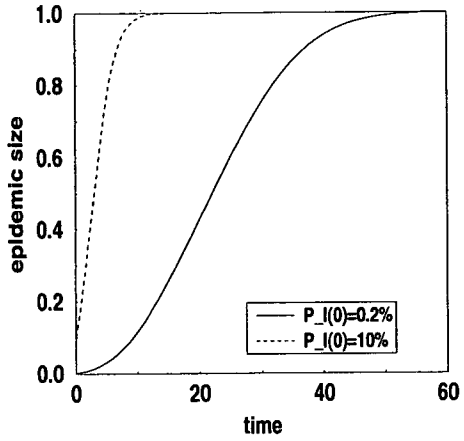


Figure 6.13: *Comparison of simulated disease progress curves with SR dispersals and different initial conditions* ($P_I(0) = 10\%$ *and* 0.2%).

Equation (2.22) can be rewritten as

$$\frac{1}{2}\frac{dP_{II}(r)}{dt} = [J_1 + J_2 f(r)]\, P_I - [J_1 + J_2 f(r) + R]\, P_{II}(r) + J_2 \sum_{\substack{s\in\mathcal{L}_{1,L/2} \\ (u(r,s,\theta)\neq 0)}} \sum_{\theta\in\Theta_s} P_{SII}(r,s,u(r,s,\theta))\, f(s), \tag{6.24}$$

where $u(r,s,\theta) = \sqrt{r^2+s^2-2rs\cos\theta}$ and $P_{SII}(r,s,u(r,s,\theta)) = P_{II}(u(r,s,\theta)) - P_{III}(r,s,u(r,s,\theta))$ (see Fig. 6.18).

The quadrats in Fig. 6.16 are identified by polar coordinates, (r_i,ϕ_i). The number of quadrats n is the same for each r_i, since all quadrats have the same angular amplitude $\phi = 2\pi/n$. The set of angular coordinates of the quadrats is denoted by $\Phi_n = \{\phi_1,\phi_2,\ldots,\phi_n\}$. As a result of the intrinsic anisotropy of the lattice, the number of lattice sites in each quadrat is not constant and varies with both r_i and ϕ_i. For a given r, on average, there are $G(r)\phi/2\pi = G(r)/n$ lattice sites in each quadrat. Fig. 6.17 illustrates this situation. At distance $r = 5$ from a given site there are $G(5) = 12$ sites on the lattice. With $n = 6$ quadrats, there are 3 lattice sites in the first quadrat, 1 in the second, and so on (see Figs 6.16 and 6.17). On average there are $12/6 = 2$ lattice sites in each quadrat. Equation (6.24) is then approximated by

$$\frac{1}{2}\frac{dP_{II}(r)}{dt} = [J_1 + J_2 f(r)]\, P_I - [J_1 + J_2 f(r) + R]\, P_{II}(r) + \frac{J_2}{n} \sum_{s\in\mathcal{L}'_{1,L/2}} \sum_{v\in\mathcal{L}_{s,s+\Delta}} \sum_{\phi\in\Phi_n} P_{SII}(r,v,u(r,v,\phi))\, f(v)G(v). \tag{6.25}$$

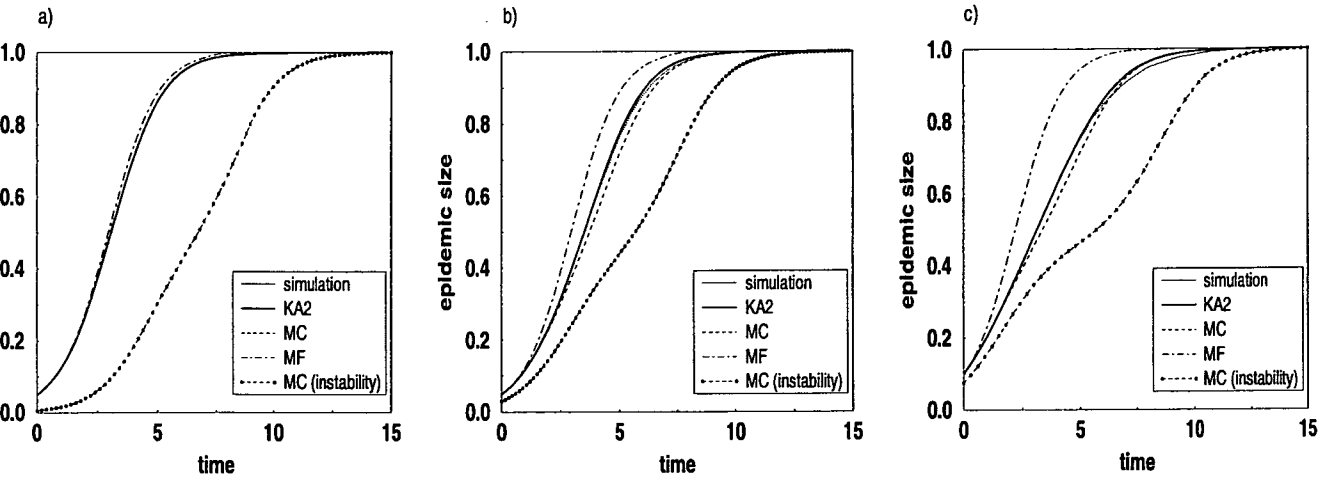


Figure 6.14: *Upper curves show the expected disease progress curves with the smallest initial infection for which MC yields a stable solution. Bottom curves show the threshold value of $P_I(0)$ for which MC exhibits numerical instabilities. a) LR (Bessel dispersal function with $\psi = 0.5$). Note that simulation, KA2 and (stable) MC curves are all indistinguishable, whilst MF overestimates P_I (upper curves); b) MR ; c) SR. In the MR and SR cases MF overestimates P_I, KA2 and simulation curves are very close (simulation visible for times $t > 5$), MC underestimates P_I.*

Assuming that $P_{SII}(r, v, u(r, v, \phi))\ f(v)$ is constant for $v \in [s, s + \Delta[$, we obtain

$$\frac{1}{2}\frac{dP_{II}(r)}{dt} = [J_1 + J_2 f(r)]\ P_I - [J_1 + J_2 f(r) + R]\ P_{II}(r) + \frac{J_2}{n} \sum_{s \in \mathcal{L}'_{1,L/2}} \sum_{\phi \in \Phi_n} P_{SII}(r, s, u(r, s, \phi))\ f(s) H(s, \Delta), \qquad (6.26)$$

which together with (6.23) form the set of ODEs for the model. Since the distances r and s belong to the discrete set $\mathcal{L}'_{1,L/2}$, equation (6.26) represents a much smaller system of equations than (2.22). The total number of ODEs in (6.23)–(6.26) is $(L/2 - 1)/\Delta + 2$. Thus, for $L = 100$ and $\Delta = 1$, we have 51 equations, compared with the 762 equations of the lattice approach for the same L.

We have chosen to let r vary between 1 and $L/2$ in (6.23) and (6.26), which is the range of distances considered in the lattice approach.

The discretisation of space into quadrats leads to the introduction of two extra parameters: Δ, the unit of discretisation of r, and n, the number of quadrats at given r (discretisation of angles). The choice of appropriate values for Δ and n is arbitrary and depends on the host distribution and the type of interactions under consideration. In this case, we wish to obtain predictions for the behaviour of a model which was simulated on a lattice. Hence we first set $\Delta = 1$, then let it vary and observe the effects of different values. The value of n is chosen in accordance

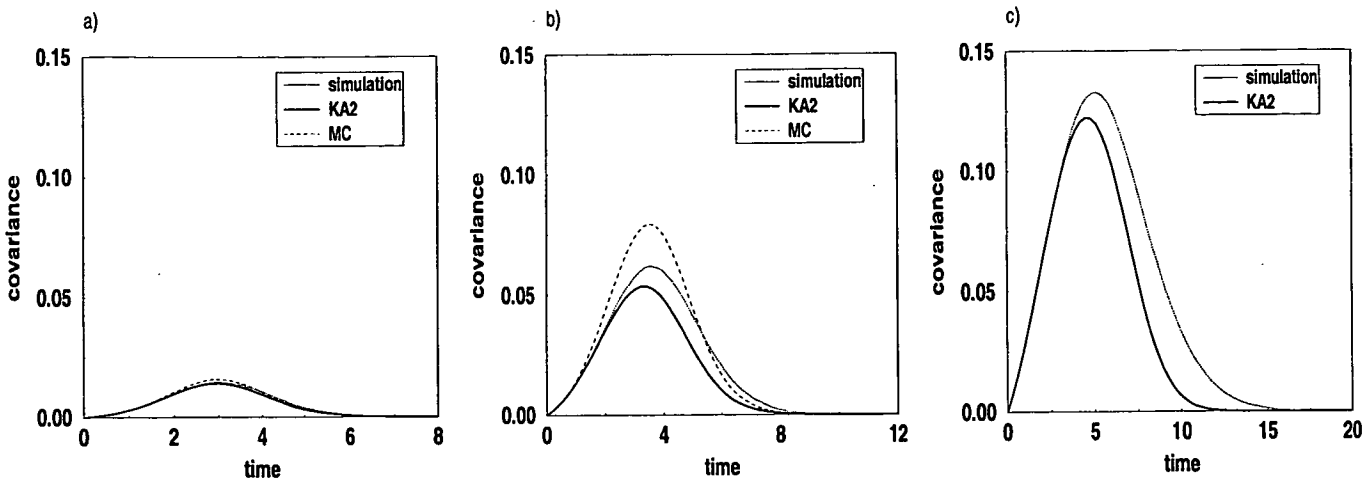


Figure 6.15: *Covariance ($\Gamma\sigma^2$) obtained from simulation and predicted by KA2 and MC. a) LR (Bessel dispersal function with $\psi = 0.5$); b) MR; c) SR (note that MC yields no solution).*

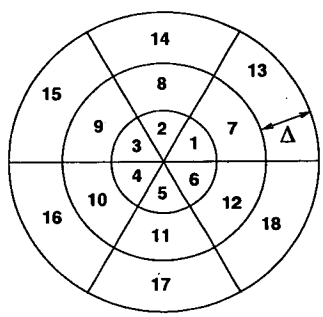


Figure 6.16: *Subdivision of continuous space in quadrats.*

with $G(r)$, which gives the number of lattice sites at distance r from a given point. Since $G(r)$ is not a continuous function of r and, on average, increases with r, n could be made dependent on r. For example, recalling that the histogram of $G(r)$ with window of width 1 behaves approximately as $2\pi r$, we could have $n(r) = 2\pi r$. However, at present we wish to avoid such a complication. For $1 \leq r \leq 50$, $G(r)$ varies between 4 and 32; we thus set $n = 16$ and then analyse the effects of different values of n on the predicted behaviour.

The contact distribution must be normalised according to the new spatial coordinates. The normalisation factor Z given by (2.4) can be rewritten as

$$Z = \sum_{r \in \mathcal{L}_{1,L}} f(r)G(r) = \sum_{s \in \mathcal{L}'_{1,L}} \sum_{r \in \mathcal{L}_{s,s+\Delta}} f(r)G(r). \tag{6.27}$$

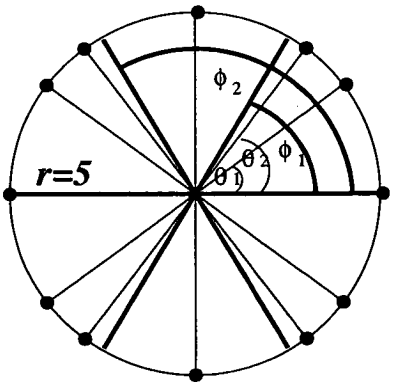


Figure 6.17: *Angular coordinates* $(\theta_1, \theta_2, \dots)$ *of lattice sites at distance* $r = 5$ *from a given site, and angular coordinates* $(\phi_1, \phi_2, \dots)$ *of quadrats. Dark dots* • *indicate lattice sites.*

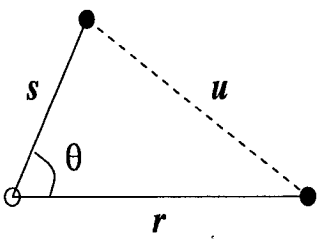


Figure 6.18: *Distances involved in equation (6.24).*

Assuming that $f(r)$ is constant in the interval $[s, s + \Delta[$ we obtain

$$Z = \sum_{s \in \mathcal{L}'_{1,L}} f(s) H(s, \Delta). \tag{6.28}$$

Equation (6.26) for $P_{II}(r)$ contains the probability $P_{II}(u(r, s, \phi))$, where $u(r, s, \phi) = \sqrt{r^2 + s^2 - 2rs\cos\phi}$. Whilst r and $s \in \mathcal{L}'_{1,L/2}$, $u(r, s, \phi)$ does not necessarily belong to the same set of distances (Fig. 6.18). If this is so then there is no equation for $P_{II}(u(r, s, \phi))$ and its value is unknown. In order to address this problem we distinguish three situations: (1) $u(r, s, \phi) < 1$; (2) $1 \leq u(r, s, \phi) \leq L/2$; (3) $u(r, s, \phi) > L/2$. In case (1) we set $P_{II}(u(r, s, \phi)) = 0$ to ensure that individuals are never at a distance smaller than 1. In case (3) we apply the MF approximation to the probability $P_{II}(u(r, s, \phi))$ (i.e. $P_{II}(u(r, s, \phi)) = P_I^2$), which is plausible since L is large. In case (2) we interpolate the function P_{II} from its known values. The values of $P_{II}(r)$ are known for all $r \in \mathcal{L}'_{1,L/2}$ and both simulation and lattice approximations have shown that P_{II} is a smooth decaying function of distance. In order to interpolate P_{II} we employ a numeric routine (E01AAF, NAG 1993)

which uses a technique of successive linear interpolations up to distances at which correlations have vanished. For larger distances we use linear interpolation.

6.5.1 Results

We now apply KA and KA2 to equation (6.26) in order to close the system (6.23)–(6.26) and obtain predictions for the expected behaviour using the continuous approach. Within the continuous approach, KA (equation 6.8) reads

$$P_{III}(r,s,u(r,s,\phi)) \approx \frac{P_{II}(r)P_{II}(s)P_{II}(u(r,s,\phi))}{P_I^3}, \tag{6.29}$$

whereas KA2 (equation 6.18) reads

$$P_{SII}(r,s,u(r,s,\phi)) \approx \frac{P_{SI}(r)P_{II}(u(r,s,\phi))P_{SI}(s)}{P_S P_I^2}. \tag{6.30}$$

6.5.1.1 Stationary behaviour

Since all approximations give very good estimates of the long–term behaviour of the LR dispersal model, we concentrate on the results for the MR and SR dispersal cases. Fig. 6.19 shows the corresponding long–term epidemic size against

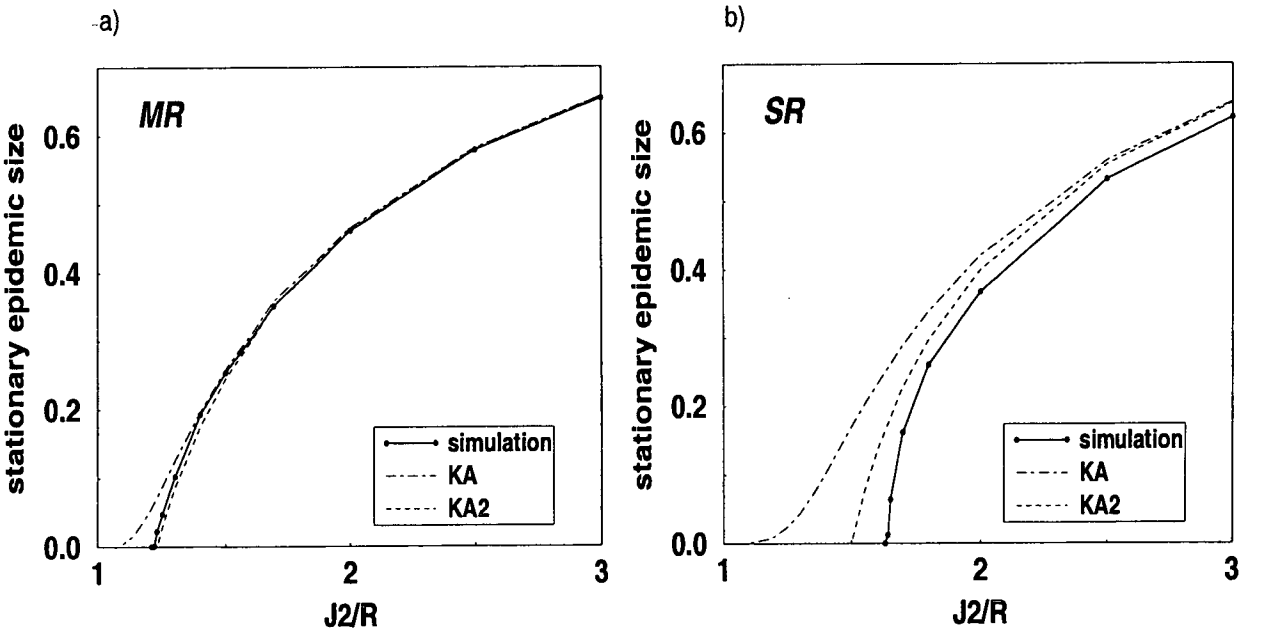


Figure 6.19: *Phase diagrams obtained from simulation and predicted by KA and KA2 in the continuous approach. a) MR; b) SR.*

infectiousness J_2/R with $J_1 = 0$, obtained from simulation of the lattice model and predicted by KA and KA2 in the continuous approach. In the MR case predictions are in very good agreement with simulation, whilst in the SR case,

as expected, the agreement is worse because in this case the effects of the lattice structure are stronger and the approximations perform less well for SR, as they do also in the lattice approach. KA overestimates the infection level, as it also does in the lattice approach, whilst KA2 slightly underestimates the infection level in the MR case and overestimates it in the SR case.

Comparison between the predictions from the lattice (Figs 6.2 and 6.3) and continuous approaches yields the following conclusions (at least for the chosen values of the quadrat parameters):

1. in the MR case, KA overestimates the epidemic size more in the lattice approach than in the continuous approach; in the SR case, KA overestimates the epidemic size more in the continuous approach than in the lattice approach;

2. KA2 systematically overestimates the epidemic size in the lattice approach, whilst in the continuous approach it overestimates the epidemic size with SR dispersal and underestimates the epidemic size with MR dispersal. This underestimation is due to an error introduced by the continuous approximation which counteracts the systematic error of KA2.

6.5.1.2 Transient behaviour

Figs 6.20 and 6.21 show the epidemic transients for three stationary disease levels

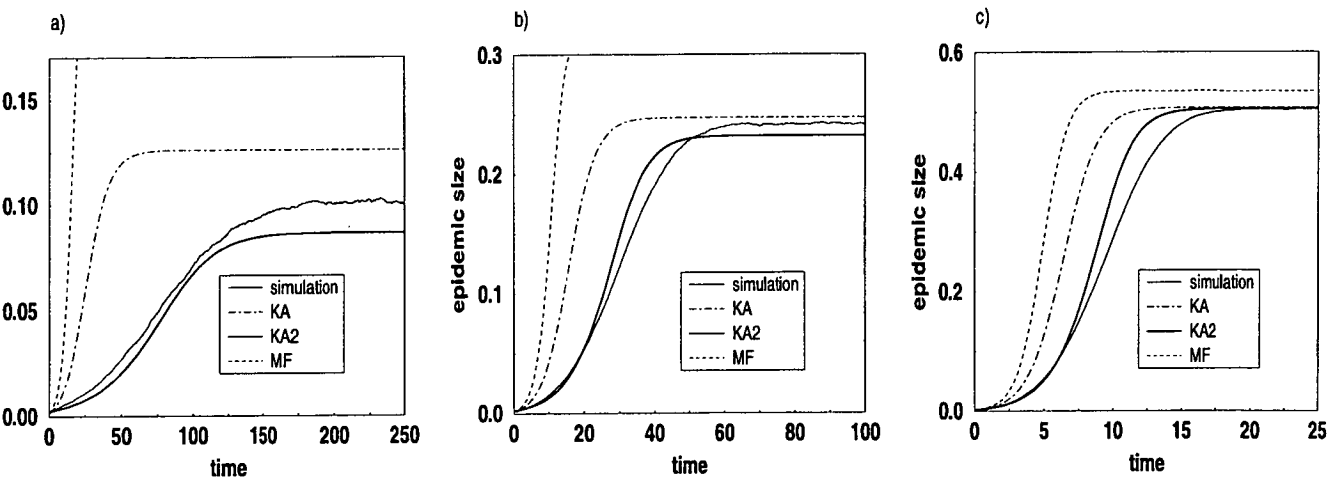


Figure 6.20: *Expected epidemic transients in the MR dispersal case. Long term epidemic size: a)* 10%, *b)* 25%, *c)* 50%. *Approximation curves from left to right (worst to best): MF, KA, KA2.*

(10%, 25%, 50%) with $J_1 = 0$ for the MR and SR dispersal cases, respectively; MF, KA and KA2 predictions are compared to simulation.

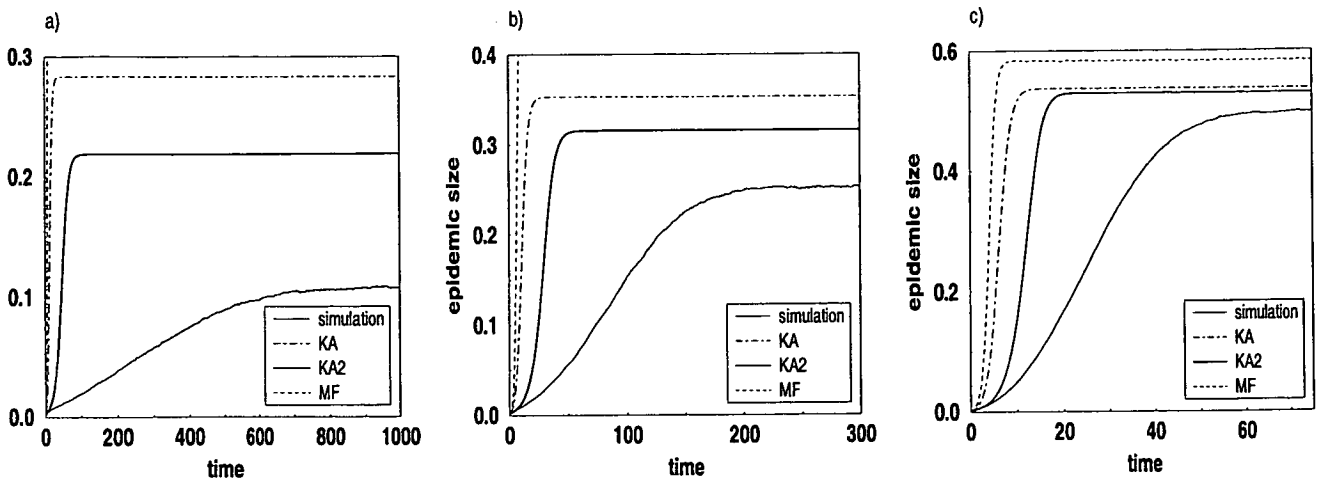


Figure 6.21: *Expected epidemic transients in the SR dispersal case. Long term epidemic size: a)* 10%, *b)* 25%, *c)* 50%. *Approximation curves from left to right (worst to best): MF, KA, KA2.*

As in the lattice approach, KA is able to capture some characteristics of the behaviour but KA2 provides better predictions. In addition, predictions are better in the MR than in the SR case. In the continuous approach the approximations predict the duration of transients better than in the lattice approach, especially with MR dispersal. In contrast to the lattice approach, KA2 slightly underestimates the disease level for some parameter values in the MR case (Figs 6.20a and b). As for the lattice approach, the improvement provided by one approximation relative to another is more modest in the transient than the stationary regime (see section 6.4.1.2).

Since the solution of the equations can be quite sensitive to variations in the discretisation parameters n and Δ, the choice of such parameters must be based on both common sense and the characteristics of the model under consideration. Here, the continuous approach has been used to predict the behaviour of a lattice–based model: simple consideration of the structure of the lattice led us to a choice of parameter values ($n = 16$ and $\Delta = 1$) which turned out to yield very good predictions.

Fig. 6.22 illustrates how KA predictions vary with n and Δ.

6.6 Conclusions

Using individual–based dynamic equations, we have developed an analytical framework for predicting the expected behaviour of spatially–extended population models with general interactions. This approach is computationally much more effi-

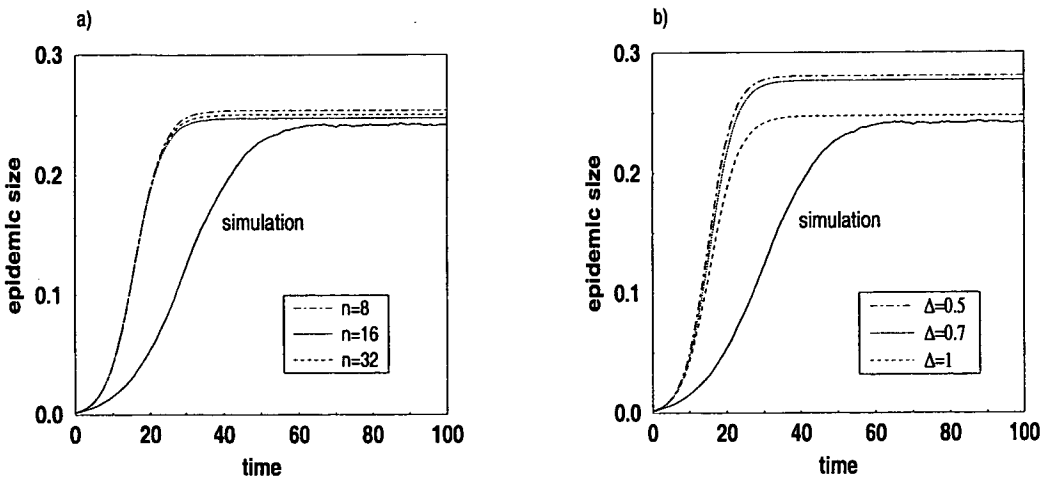


Figure 6.22: *Variation of the theoretical predictions with parameters n (a) and Δ (b). Simulated epidemic corresponds to MR dispersal. a) Δ = 1; b) n = 16.*

cient than the stochastic simulation of the model and therefore provides a useful alternative tool for exploring the behaviour. Moreover, since it is accurate over wide and relevant parameter regions, the approach might also be used for fitting observations and estimating basic parameters.

The study was conducted in the context of the plant epidemic model introduced in chapter 2. The dynamics of the model are described by a system of ODEs which can only be solved approximately using closure relations. Solutions were tested against stochastic simulation. We found that the moment closure approximation (MC) proposed by Bolker & Pacala (1997, 1999) is inadequate when the spatial correlations in the disease pattern are not small. We then proposed and tested alternative closure schemes which proved to be more robust and provide better descriptions of the model behaviour. In particular, we demonstrated that the modified Kirkwood approximation (KA2) is superior to all the other approximations which we considered and is able to offer very accurate predictions for both the epidemic size and the spatial correlations over most of the parameter space. The approximations perform best for LR and MR interactions, whereas they are less effective in the SR case; however, even for SR interaction models, KA2 outperforms most of the existing cluster approximations. A common feature of both types of closure approximations is their greater ability to predict accurately stationary behaviour rather than transient behaviour.

The system of dynamic equations was solved within lattice and continuous formulations. In the lattice approach the equations describe exactly the expected simulated dynamics, thus the errors introduced by closure approximations do not mix with errors resulting from continuous–media approximations. The continuous approach was devised in order to reduce computational cost and extend the

method to models defined in continuous space. It is based on the discretisation of continuous space into a mosaic of quadrats within which the densities are assumed to be constant. This leads to a significant decrease in the number of equations.

In this study we have focussed on the lattice approach because it represents the most appropriate choice for predicting the behaviour of a lattice–based model and because the primary objective was to test the closure approximations as accurately as possible. Unsurprisingly, we found that the continuous approach is less accurate in the SR case where the lattice structure is more important. Nevertheless, the results obtained within this approach are remarkably good in some parameter regions.

This work can be further developed in various ways. Two examples would be:

1. the development of a mixed approach that takes into account interactions between individuals (as in the lattice approach) up to a given distance and beyond this distance uses quadrats (as in the continuous approach); such an approach would combine the benefits of the lattice and continuous formulations: a smaller number of equations than with the lattice approach and more accurate predictions than with the continuous approach;

2. the extension of the method to systems which are spatially non–stationary. Possible causes of non–stationarity include an irregular spatial distribution of hosts, the effect of boundaries on the interaction neighbourhoods of different hosts (Filipe *et al.*, 2000), spatially heterogeneous primary infection (Gilligan, 1985), and specific initial conditions. Such an extension should improve the ability of the approximations to capture transient behaviour.

Chapter 7

Heterogeneous mixing models

7.1 Introduction

In classical epidemiological models (see for example Bailey, 1975), the incidence rate, defined as the rate at which new infections occur in a population, is represented by a term bilinear in the number (or fraction) of susceptibles and infecteds. This assumption, implying that susceptibles and infecteds mix homogeneously in space, is plausible when dealing with small groups of individuals such as families but often inadequate for larger and spatially extended communities. In particular, the development of epidemics in plant populations, whose individuals are unable to move freely, is strongly affected by the dynamically generated spatial heterogeneity in disease pattern. When the assumption of homogeneous mixing is inappropriate, it would be preferable to introduce the necessary population structure in the model and represent heterogeneous mixing directly (Liu *et al.*, 1987). However, fitting spatially–explicit models to field observations is a difficult task because it entails the collection of large samples of temporal and spatial data.

Spatial heterogeneity and other biological factors have motivated a modification of the standard bilinear form of the incidence rate, proportional to IS (where I is the number of infectives and S the number of susceptibles). Departures from this model were first considered by Severo (1969), who proposed a model in the context of human epidemiology in which the incidence rate was proportional to $I^a S^{1-b}$. His choice was motivated by the following arguments.

1. Epidemics which develop very slowly even when the number of infectives is large can be described by an incidence rate essentially independent of the number of infectives, i.e. by small values of a. This might be the case, for example, when the rate of physical contact in a community is high but the contagion is determined by the susceptibility of healthy individuals rather than by the infectiousness of infectives.

2. Fast–developing epidemics, in which nonlinear effects make the presence of k infectives more infectious than k times the infectiousness of one infective, can be described by large values of a. An example would be epidemics where there is a positive co–operativity of infectives.

3. The existence of large numbers of susceptibles might lower the transmission rate of the disease: if individuals can only transmit the infection at close contact, a high population density might inhibit the ability of an infective to move around the community. A small value of b would then imply that there is greater safety in having many rather than fewer susceptibles, whilst b close to unity would mean that the chance of a given susceptible becoming infected is almost independent of the total number of susceptibles.

It has been shown (Liu *et al.*, 1986; Liu *et al.*, 1987; Hethcote & van den Driessche, 1991) that the qualitative dynamic behaviour of SEIR and SIRS models with nonlinear incidence rates can be very different from that of models with traditional bilinear incidence. Whilst the latter exhibit a simple pattern of behaviour, with epidemic extinction below - and a steady state above - a certain transmission threshold, models with nonlinear incidence rates can have multiple equilibria and periodic solutions which may have epidemiological significance.

Recently, Gubbins & Gilligan (1997 a, b) analysed the role of spatial heterogeneity in the persistence of a lettuce fungal disease in a host-parasite-hyperparasite system where a fungal hyperparasite was used as a biocontrol. Spatial heterogeneity was incorporated in their model by relaxing the assumption of homogeneous mixing of the populations, i.e. by assuming the response of the hyperparasite to changes in parasite density to be nonlinear. The model was fitted to experimental data and then used for prediction. It was shown that whilst the homogeneous mixing model predicted parasite extinction, the heterogeneous mixing model prevented fade–out and considerably improved the fit of the model. The authors identified various mechanisms which could be responsible for the nonlinear functional response. However, since there was no experimental evidence in favour of either hypothesis, they concluded that the nonlinearity probably arose from spatially–heterogeneous mixing in the population (Gubbins & Gilligan, 1997 a).

Whilst heterogeneous mixing models might describe effectively the spatial heterogeneity which naturally arises in real systems, it remains to be seen how the additional mixing parameters relate to the spatial structure. We simulate stochastic spatially–explicit models for plant epidemics in order to relate the parameters of the heterogeneous mixing model to the self–generated spatial structure in the distribution of the disease. We shall show that the heterogeneous mixing model

provides an excellent fit to the simulated data in the transient but not in the stationary regime, which might relate to differences in the spatial patterns of disease in the two stages. The fitted parameters are constant in time but vary with the basic parameters of the stochastic model (dispersal exponent and infectiousness).

7.2 Spatially–heterogeneous mixing model

In chapter 2 we introduced a spatio–temporal model for the spread of plant epidemics and derived a system of deterministic differential equations for describing its dynamics. We used the function $\Gamma(t)$ (2.31), a convolution of the spatial autocorrelation function and the contact distribution, to write the first equation of the system (2.21)–(2.22) in the form (2.30). This formulation shows that the actual incidence rate is modified by the presence of Γ. The standard bilinear incidence rate, $P_I(1 - P_I)$, is valid only when $\Gamma = 0$ and the model behaves like the Mean Field (MF) approximation (chapter 3; Filipe *et al.*, 2000). When correlations are large or long-ranged, for example in the proximity of the persistence threshold, the deviation from the MF behaviour and from the classical bilinear incidence rates can be substantial.

In the light of the parameterised heterogeneous mixing model introduced by Liu *et al.* (1987) and considered by Gubbins & Gilligan (1997 a, b), the incidence rate $\mathcal{I}$, corresponding to the epidemic model considered here, is given by

$$\mathcal{I}(t) = k\ P_I(t)^{\alpha}\ [1 - P_I(t)]^{\beta}, \tag{7.1}$$

where k, α and β are real and positive fitting parameters. The kinetic equation for the time evolution of $P_I(t)$ is therefore given by

$$\frac{dP_I(t)}{dt} = -R\ P_I(t) + J_2\ \mathcal{I}(t). \tag{7.2}$$

From (2.30) and (7.2) we obtain an expression for Γ in terms of the expected fraction of infecteds and the fitting parameters

$$\Gamma(t) = 1 - k\ P_I(t)^{\alpha-1}\ [1 - P_I(t)]^{\beta-1}. \tag{7.3}$$

This expression relates the empirical parameters k, α and β to the spatial correlations.

7.3 Simulation results

We now wish to test the heterogeneous mixing model (7.2) against simulation of the spatially–extended model. This can be done by finding the parameter values which best fit the relation (7.3) over the whole epidemic.

Stochastic realisations of the spatially explicit model are realised on a lattice of linear size $L = 100$. Epidemics are initiated with 20 randomly distributed infecteds in the cases of LR, MR and SR dispersals. The incidence rate

$$\mathcal{I}_{sim}(t) = P_I(t)[1 - P_I(t)][1 - \Gamma(t)] \tag{7.4}$$

is estimated at given points in time using averages of $P_I(t)$ and $\Gamma(t)$ over a large number of realisations (100 for the SI model, and 20 for the SIS model).

7.3.1 SI model

Fig. 7.1 exhibits the simulation incidence rate $\mathcal{I}_{sim}$ (7.4) and the fitted curves $\mathcal{I}$ (7.1), corresponding to the incidence rate of the heterogeneous model, against the average fraction of infecteds P_I for the three dispersal cases. The curve corresponding to the MF model ($\Gamma = 0$) is also shown. The model fits the simulation

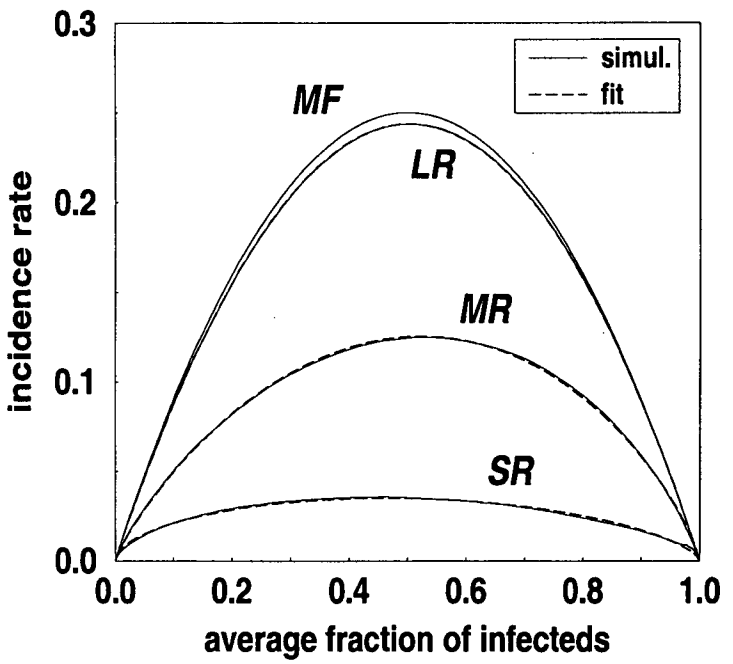


Figure 7.1: *Simulation incidence rates and fitted curves (7.1) corresponding to the heterogeneous mixing model vs $P_I(t)$ for different dispersal cases (LR, MR and SR from top to bottom). Also shown: MF bilinear term (top).*

very well, especially for LR and MR, showing that it is capable of capturing the behaviour of the spatially explicit model throughout the progress of the epidemic. In the SR case, the fit is slightly poorer because of the large stochastic fluctuations typical of this parameter region. In principle, larger samples of realisations could lead to a smoother curve $\mathcal{I}_{sim}$ and hence a better model fit. A measure of

the variability of the incidence rate in the SR dispersal case is illustrated in Fig. 7.2 where $\mathcal{I}_{sim}$ and the curves

$$\mathcal{I}^{\pm}_{sim}(t) = P_I(t)\ [1 - P_I(t)]\ [1 - \Gamma(t) \pm \sigma[\Gamma(t)]] \tag{7.5}$$

are plotted against the average fraction of infecteds.

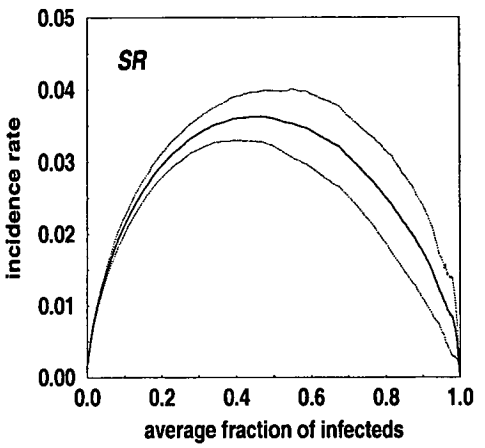


Figure 7.2: *Variability of the incidence rate in the SR dispersal case. Sample consists of 100 realisations.*

The values of the fitted parameters k, α and β, listed in table 7.1, vary with the type of dispersal in the expected way. The value of the fitted parameters is

dispersal range	k	α	β
∞ (MF)	1	1	1
LR	0.967	0.999	0.989
MR	0.394	0.863	0.790
SR	0.079	0.546	0.635

Table 7.1: *Values of the fitted parameters of the heterogeneous mixing model for different dispersal cases.*

very close to 1 in the LR case and decreases as the range of dispersal decreases. In the SR case it is minimum.

An argument suggested by McNeil (1972) and reported by Bailey (1975) sheds some light on the meaning of the values of the parameters in the SI model. Consider an epidemic which spreads deterministically from an infected circular area of radius r. The fraction of infecteds is proportional to the area of the circle ($P_I \propto \pi r^2$) and, in the SR case, the rate of spread is proportional to the length

of the boundary, i.e. the square root of the fraction of infecteds

$$\frac{dP_I}{dt} \propto \frac{d\pi r^2}{dt} = 2\pi r \propto P_I^{1/2}. \tag{7.6}$$

This result agrees well with the estimated value of α for the SR case ($\alpha = 0.546$, see table 7.1). The diffusion process described above can in fact be seen as an approximation of an SI model with SR dispersal: an epidemic with no recovery, initiated by one diseased individual which can only infect its neighbouring sites, generates infected regions which are approximately circular. Hence, the incidence rate does not depend on the density of infecteds inside the cluster but on the density of infecteds at the boundary, where susceptible–infected contacts occur. The parameter β is also expected to decrease with the range of dispersal: the incidence rate cannot depend linearly on the fraction of susceptibles since infecteds are aggregated, hence the number of possible susceptible–infected contacts is limited.

We now substitute the values of the fitted parameters in model (7.1) into the differential equation (7.2) and solve it numerically. Fig. 7.3 compares the

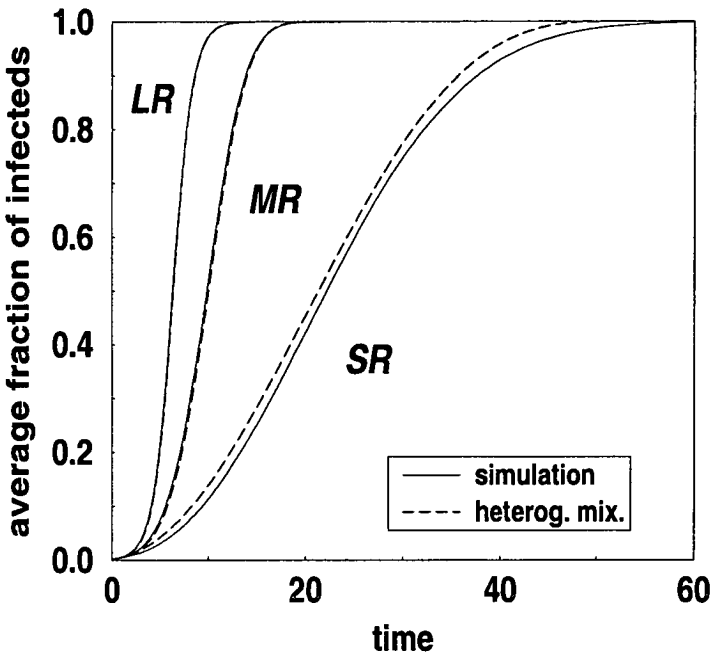


Figure 7.3: *Disease progress curves for different dispersals. Simulation and heterogeneous mixing model curves.*

disease progress curves obtained from simulation with the predictions of the parameterised heterogeneous mixing model in the three dispersal cases. Model and simulation are very close, especially for LR and MR dispersal. Unsurprisingly, in the SR case the prediction slightly overestimates the disease level.

7.3.2 SIS model

In an infinite system with recovery of infecteds, the epidemic reaches a stationary state where the fraction of infecteds is constant in time and given by (2.33). Hence the incidence rate is also constant and given by $\mathcal{I} = RP_I/J_2$. In a finite system, instead, the epidemic reaches a long–term quasi–stationary state in which the fraction of infecteds fluctuates around the equilibrium value (2.33). As a consequence, the incidence rate also fluctuates, exhibiting behaviour different from that in the transient. In order to fit the heterogeneous mixing model to simulation, we exclude the stationary regime after arbitrarily defining the end of the transient. Moreover, the density of points in the plot of $\mathcal{I}_{sim}$ vs P_I is greater in the stationary regime. In order to avoid giving too much weight in the least square fit to this part of the curve, we smooth the data using running averages.

The simulation curves presented in this section represent averages over 20 stochastic realisations of the model. Samples are smaller in this case because simulation is more time–consuming and stochastic fluctuations are smaller for the SIS than the SI model.

We consider two different stationary epidemic sizes (50% and 72% infecteds) in each of the three dispersal cases. Fig. 7.4 shows simulation and fitted incidence rate curves against the average fraction of infecteds. The heterogeneous mixing

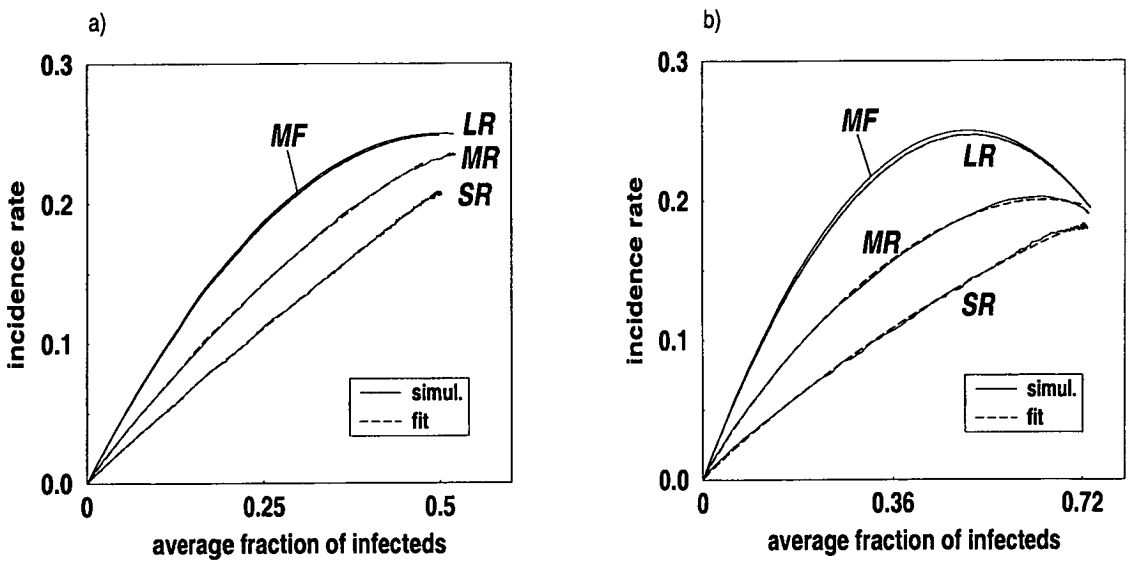


Figure 7.4: *Simulation incidence rates and fitted curves (7.1) corresponding to the heterogeneous mixing model vs $P_I(t)$ for different dispersal cases (LR, MR and SR from top to bottom). Also shown: MF bilinear term (top). a) $P_I(\infty) = 50\%$; b) $P_I(\infty) = 72\%$.*

model (7.1) provides a good fit to the incidence rate of the stochastic model for both stationary states in all dispersal cases, although fits are more accurate in the case of 50% infecteds.

Table 7.2 shows the values of the fitted parameters k, α and β corresponding to Fig. 7.4. As for the SI model, in the LR case the fitted parameters are close

dispersal range	$(P_I)_\infty$	k	α	β
∞ (MF)	50%	1	1	1
	72%	1	1	1
LR	50%	0.988	1.002	0.987
	72%	0.979	1.002	0.984
MR	50%	0.608	0.961	0.414
	72%	0.513	0.926	0.518
SR	50%	0.428	0.959	0.072
	72%	0.289	0.908	0.123

Table 7.2: *Values of the fitted parameters of the heterogeneous mixing model for different dispersals and stationary epidemic sizes.*

to 1 and decrease with the range of dispersal. Comparing the two asymptotic epidemic sizes, we observe that with 72% infecteds:

1. k and α are smaller than with 50% infecteds;

2. β is larger than with 50% infecteds.

Comparing the SI and the SIS models, we observe that:

1. k and α decrease less with decreasing dispersal range in the SIS model than in the SI model;

2. β decreases more with decreasing dispersal range in the SIS model than in the SI model.

As before, we use the fitted parameters to solve equation (7.2) and obtain the corresponding disease progress curves. These are compared to simulation in Fig. 7.5, which shows that the heterogeneous model gives reasonably good predictions except for the SR case. As expected, the model curves fail to predict the long–term behaviour of the epidemic, with the exception of the LR case, which is very close to MF. In the MR and SR cases, in contrast to the SI case, the model systematically underestimates the disease level.

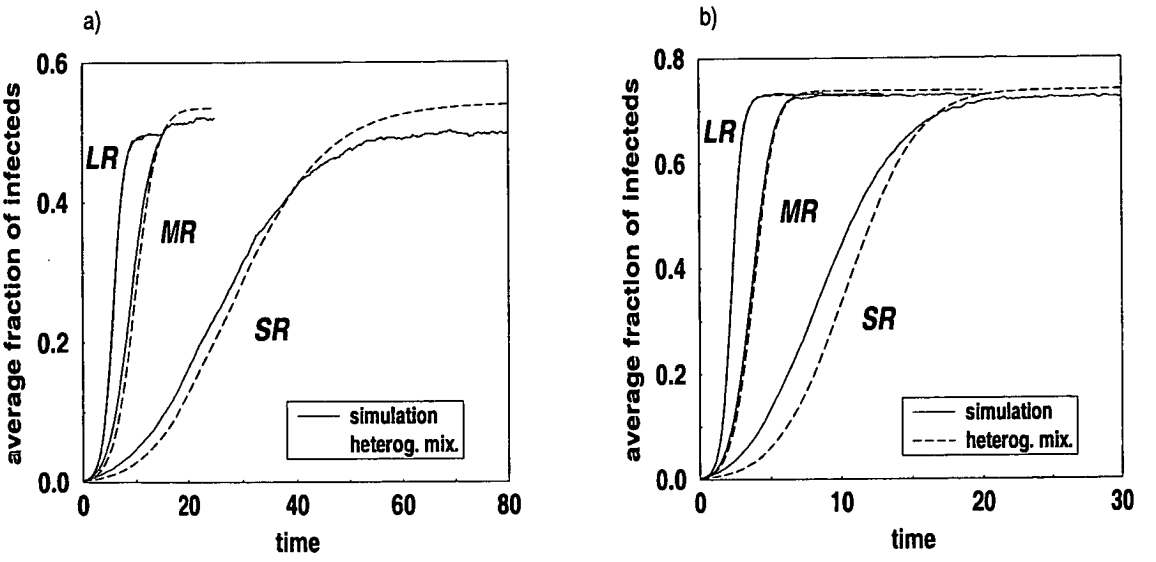


Figure 7.5: *Disease progress curves for different dispersals. Simulation and heterogeneous mixing model curves. a)* $P_I(\infty) = 50\%$*; b)* $P_I(\infty) = 72\%$.

7.4 Conclusions

In this chapter we have used simulation of stochastic, spatially–explicit models to test empirical extensions of deterministic, non–spatial models which attempt to account for heterogeneous mixing of susceptibles and infecteds. In order to relax the assumption of homogeneous spatial mixing, these models consider incidence rates which are nonlinear in the number (or fraction) of susceptibles and infecteds. It has been shown that such models exhibit a much wider range of dynamic behaviour (Liu *et al.*, 1986; Liu *et al.*, 1987; Hethcote & van den Driessche, 1991) and provide a considerably better fit to observation data (Gubbins & Gilligan, 1997 a, b) than models with bilinear incidence rates. We have examined the relation between the empirical parameters of the heterogeneous mixing model and the correlations in the self–generated spatial structure of the disease.

We have found that the heterogeneous mixing model describes the incidence rate of the spatial model very well throughout the epidemic in the absence of recovery (SI model), and in the transient regime in the presence of recovery (SIS model).

Epidemics with different stationary disease levels are characterised by different incidence rate curves. Different epidemic sizes, in fact, correspond to different values of the infectiousness J_2 and therefore to different spatial distributions of the infecteds.

At least three empirical parameters are required for describing the shape of the incidence rate curve of the full model: two parameters describing the slope at the initial and final stages of the epidemic, corresponding to $P_I(0)$ and $P_I(t_{max})$, and one parameter describing the location of the maximum of the curve. The fact that only three parameters are needed to obtain a good fit to simulation seems to us striking, especially if we consider the relation (7.3) between α, β, k and $\Gamma(t)$ and $P_I(t)$.

The fit of the heterogeneous model to simulation is particularly good for the LR dispersal case: the empirical parameters are very close to 1 and the incidence rate differs little from the bilinear incidence rate typical of MF models. The values of the fitted parameters decrease with decreasing range of interaction and are minimum for SR dispersal. In the SI case, epidemics with SR dispersal can be approximated by diffusion processes where the incidence rate is proportional to the square root of the fraction of infecteds. In the SIS case, the incidence rate curves become almost linear and β becomes very small for SR dispersal, suggesting that the incidence rate is only slighlty dependent on the fraction of susceptibles. Results for the MR case are intermediate.

The values of the fitted parameters were used to obtain predictions for the disease progress curves, which were then compared with simulation. The heterogeneous mixing model provides good predictions for all dispersal cases, especially LR and MR, in the absence of a recovery process (SI model). In the SIS model, predictions are very good for the LR case in both the transient and the stationary regime, and for the MR case in the transient regime. Predictions are less good for SR dispersal.

This study suggests that models with nonlinear incidence rates are suitable for describing the dynamics of spatially–extended SIS and SI models, provided that they comprise at least three empirical parameters. We showed that the dependence of the new parameters on the basic parameters of the spatial model (the range of dispersal and the infectiousness) takes place through their relation with spatial correlations.

Chapter 8

Summary

The main topic of this research project has been the investigation of the spatial components of plant disease epidemics. Although for many years the emphasis of research has been on the temporal aspects of plant epidemics (Vanderplank, 1963; Leonard & Fry, 1986; Campbell & Madden, 1990), interest in the effects of space has grown enormously in the past decade (Jeger, 1989; Durrett & Levin, 1994 a; Russell, 1996; Filipe *at al.*, 2000). Great effort has been made to demonstrate the limitations and inadequacies of a purely temporal approach to plant disease epidemics, whilst a spatial perspective has offered new insights and opened new areas of research.

Various developments have enhanced the ability of modellers to tackle difficult spatial problems, including the formulation of analytical methods (e.g. Filipe & Gibson, 1998, 2000; Bolker & Pacala, 1997) and the advent of more powerful computers which enable the investigation of behaviour through the simulation of complex spatial models (e.g. Shaw, 1995; Filipe *at al.*, 2000; Filipe & Maule, 2000 a). At present, however, current theory suffers from a lack of methodology capable of dealing with spatial and stochastic aspects of model behaviour in a realistic way, and many important questions still remain unanswered.

Such considerations regard not only the study of plant epidemiology but all the disciplines in the broad field of population biology (Renshaw, 1991). The study of the spatial aspects of population dynamics is the subject of the theory of biological invasions (Hengeveld, 1989). Invasions can be studied from widely different viewpoints and with different aims, as well as at different spatio–temporal scales. In ecology, an invasion corresponds to the migration of a species into a territory previously not colonised by that species; instead, in population genetics it corresponds to the spatial progression of qualitatively advantageous traits; in epidemiology it concerns the spread of diseases; and in biogeography it relates to the effect of large–scale phenomena (including species migration, epidemic outbreaks and spread of genetic material) on broad geographical areas. In principle,

progress in the development of techniques and methodologies for the study of spatial aspects in any of these disciplines can be fruitfully extended to all the others.

In this thesis we have attempted to address some of the relevant questions in the current theory of population dynamics using a simple model for plant epidemics. A simple model that comprises important characteristics such as spatial extension and stochasticity represents a convenient framework within which mathematical methods for the study of spatial dynamics can be developed and tested (chapter 2). The generality of the model considered ensures that the results of our investigation and the mathematical approaches developed can be extended to a wide range of spatio–temporal models.

We have investigated the effects of different mechanisms of pathogen dispersals on the establishment, development and persistence of plant disease epidemics. In order to understand the inter–relation between spatial structure and temporal dynamics of the process we adopted a range of techniques which include simulation and analytical approaches. The analysis of the nonlinear model considered is very difficult because it involves an infinite hierarchy of ordinary differential equations which is not solvable (chapter 2). It is here that simulation becomes a valuable tool for the study of the model behaviour (chapters 3, 5 and 7). Yet however useful simulation may be, an analytical solution is always preferable (Morgan, 1984). Simulation approaches can be both time–consuming and expensive; above all they are, by definition, only concerned with particular models, and hence their results are unsuitable to be generalised or used in direct quantitative prediction (Jeger, 1986). Thus, approximate analytical solutions are sought; and simulation provides a check on the accuracy of the assumptions made (Morgan, 1984; Filipe & Maule, 2000 b; chapters 4 and 6). It is precisely this engagement between simulation and analytical methods that allows the potential of each approach to be explored and exploited in a profitable way (Renshaw, 1991). This is the procedure we followed in order to accomplish the research project presented in this thesis.

Summary of results and possible developments

First of all we considered a general stochastic model in which individuals interact according to power–law contact distributions (Mollison, 1977). Using stochastic simulation, we addressed a number of fundamental questions regarding the size, variability and persistence of epidemics, and in particular their relation to the type of dispersal process, the presence of spatial boundaries, the size of the system,

and the initial conditions (Filipe *et al.* 2000; Filipe & Maule, 2000 a). Special attention was paid to the structure and dynamics of spatial correlations and to their relation to the expectation and variability of epidemic size (chapter 3).

As an alternative to simulation, we then proposed a general approach for predicting the behaviour of the model (chapters 2 and 6) built on individual–based ordinary differential equations and various closure approximations. The approach is mathematically simple, robust and capable of dealing with and predicting the development of spatial correlations. Its flexibility and clarity make it suitable for generalisations and applicable to a wide range of interacting particle models. Analytical predictions were tested against simulation and found to be very accurate in most of the parameter space. Closure approximations performed less well in parameter regions where spatial correlations are most important. However, even in particularly unfavourable conditions, the approach outperformed existing analytical methods (Filipe & Maule, 2000 b, c), such as moment closure techniques (Bolker & Pacala, 1997; Bolker, 1999) and, in the case of short–range interactions, existing cluster approximations (pairwise and squarewise approximations, Filipe & Gibson, 1998, 2000; see also chapter 4). The approach was formulated in two versions: a lattice approach, designed for discrete distributions of individuals, and a "continuous" approach, in which discrete individuals are located in continuous space or in which there is a continuous mass–density of individuals. Various, potentially–interesting developments of this work are possible, and some examples are:

- the extension of the method to spatially non–stationary systems: we would expect this to improve the ability of the approximations to capture transient behaviour;

- the refinement of the continuous approach through a deeper understanding of its dependence on the space–discretisation parameters: this could lead to less arbitrary choices of parameter values;

- the elaboration of a hybrid approach mixing the lattice (for short distances) and continuous (for large distances) approaches;

- the exploration and development of other closure approximations which could better capture behaviour in regions where correlations are large.

- the extension of the method to more general models which include a latent period and/or the acquisition of temporary or permanent immunity (SEIS, SEIR, SEIRS models).

Analytical methods based on cluster approximations are commonly used for predicting the dynamics of stochastic models characterised by nearest–neighbour interactions (Sato *et al.*, 1994; Filipe & Gibson, 1998, 2000), which are a particular case of the general epidemic model considered here. In chapter 4 we reviewed existing approximation techniques and compared their predictions with simulation of the stochastic process. We then formulated a nearest–neighbour model for spatially–anisotropic disease spread where the rates of disease transmission along the horizontal or vertical axes of the lattice can be different. The effects of various degrees of strength of the spatial anisotropy on the size of epidemics and the disease distribution were investigated using simulation. Cluster approximations were then extended to the model and their ability to capture the effects of spatial anisotropy was tested against simulation (chapter 5). We found that the performance of the approximations worsens as the anisotropic effects increase and only the sophisticated hybrid pairwise approximation (Filipe, 1999 a, b) captures the behaviour of the model satisfactorily. More general models, which assume different rates of infection in all four directions and are able to account for different causes of spatial anisotropy (for example, the effect of the wind in the spread of airborne propagules), could be formulated and investigated in a similar way. The closure schemes developed for models with general interactions could also be applied to the anisotropic model.

Simulation was also used to test empirical extensions of non–spatial deterministic models (Liu *et al.*, 1987; Gubbins & Gilligan, 1997 a, b) which attempt to account for spatially heterogeneous mixing of susceptibles and infecteds (chapter 7). We derived a relation between the mixing parameters of the models and the self–generated spatial structure in the distribution of disease involving spatial correlations. Further studies would be needed in order to understand better the dependence of the mixing parameters on the parameters of the spatial stochastic model. This study suggests that these are plausible models for implicit account of heterogeneous mixing provided that they have at least three parameters.

In this thesis different approaches were developed and their potential for predicting the behaviour of spatial stochastic models was demonstrated. The methods presented are not restricted to the specific models to which they were applied but could provide a general framework for studying aspects of a large variety of IPS models.

References

1. Aylor, D. E., 1987. Deposition gradients of urediniospores of *Puccinia recondita* near a source. *Phytopathol.* **77**, 1442.

2. Aylor, D. E. and Ferrandino, F. J., 1989. Temporal and spatial development of bean rust epidemics initiated from an inoculated line source. *Phytopathol.* **79**, 146.

3. Aylor, D. E. and Ferrandino, F. J., 1990. Initial spread of bean rust close to an inoculated bean leaf. *Phytopathol.* **80**, 1469.

4. Bailey, N. T. J., 1975. The mathematical theory of infectious diseases and its applications. London: Griffin.

5. Bassler, K. E. and Browne, D. A., 1996. Three species monomer-monomer model: a mean-field analysis and Monte Carlo study. *Phys. Rev. E* **55**, 5225.

6. Bassler, K. E. and Browne, D. A., 1998. The three species monomer-monomer model in the reaction-controlled limit. *J. Phys. A* **31**, 6309.

7. ben-Avraham, D. and Köhler, J., 1992. Mean-field (N,M)-cluster approximation for lattice models. *Phys. Rev. A* **45**, 8358.

8. Bethe, H. A., 1935. Statistical theory of superlattices. *Proc. Roy. Soc. A* **150**, 552.

9. Binder, K. and Heermann, D. W., 1992. Monte Carlo simulation in Statistical Physics. Berlin; London: Springer–Verlag.

10. Bolker, B. M., 1999. Analytic models for the patchy spread of plant disease. *Bull. Math. Biol.* **61**, 849.

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

12. Bolker, B. M. and Pacala, S. W., 1999. Spatial moment equations for plant competition: understanding spatial strategies and the advantages of short dispersal. *American Nat.* **153**, 575.

13. Campbell, C. L. and Madden, L. V., 1990. Introduction to plant disease epidemiology. New York: Wiley.

14. Cannas, S. A., 1998. Phase diagram of a stochastic cellular automaton with long–range interactions. *Physica A* **258**, 32.

15. Cannas, S. A. and Tamarit, F. A., 1996. Long–range interactions and nonextensivity in ferromagnet spin models. *Phys. Rev. B* **54**, R12661.

16. Chesson, P. L., 1981. Models for spatially distributed populations: the effect of within–patch variability. *Theor. Pop. Biol.* **19**, 288.

17. Cliff, A. D. and Ord, J. K., 1981. Spatial processes: models & applications. London: Pion.

18. Czárán, T., 1998. Spatiotemporal models of population and community dynamics. London: Chapman & Hall.

19. de Jong, M. C. M., Diekmann, O., Heesterbeek, H., 1995. How does transmission of infection depend on population size? In: Epidemic models and their relation to data. Ed. Mollison, D. Cambridge University Press.

20. Dickman, R., 1986. Kinetic phase transition in a surface-reaction model: mean-field theory. *Phys. Rev. A* **34**, 4246.

21. Durrett, R. and Levin, S. A., 1994 a. The importance of being discrete (and spatial). *Theor. Pop. Biol.* **46**, 363.

22. Durrett, R. and Levin, S. A., 1994 b. Stochastic spatial models: a user's guide to ecological applications. *Phil. Trans. R. Soc. Lond. B* **343**, 329.

23. EPCC, 1995. *MPI: A Message-Passing Interface Standard.* Document available at
`http://www.epcc.ed.ac.uk/epcc-tec/documents/mpi-11-html/`.

24. Ferrandino, F., 1996. Two–dimensional distance class analysis of disease–incidence data: problems and possible solutions. *Phytopathol.* **86**, 685.

25. Ferrandino, F., 1998. Past non–randomness and aggregation to spatial correlation: 2DCORR, a new approach for discrete data. *Phytopathol.* **88**, 84.

26. Filipe, J. A. N., 1999 a. Hybrid closure-approximation to epidemic models. *Physica A* **266**, 238.

27. Filipe, J. A. N., 1999 b. Hybrid approximation to a generalised Contact Process. Unpublished manuscript.

28. Filipe, J. A. N. and Gibson, G. J., 1998. Studying and approximating spatio–temporal models for epidemic spread and control. *Phil. Trans. R. Soc. Lond. B* **343**, 2153.

29. Filipe, J. A. N. and Gibson, G. J., 2000. Comparing approximations to spatio–temporal models for epidemics with local spread. Accepted.

30. Filipe, J. A. N., Maule, M. M., Hambly, B. M., 2000. Effects of dispersal mechanisms on spatio-temporal development of epidemics. I. Stationary behaviour. Submitted.

31. Filipe, J. A. N., Maule, M. M., 2000 a. Effects of dispersal mechanisms on spatio-temporal development of epidemics. II. Transient behaviour. Submitted.

32. Filipe, J. A. N. and Maule, M. M., 2000 b. Analytical methods for predicting the spatio-temporal behaviour of stochastic population models with general spatial interactions. Accepted.

33. Filipe, J. A. N. and Maule, M. M., 2000 c. Comment on "Analytic models for the patchy spread of plant disease". Submitted.

34. Filipe, J. A. N. and Rodgers, G. J., 1995. Theoretical and numerical studies of chemisorption on a line with precursor layer diffusion. *Phys. Rev. E* **52**, 6044.

35. Fitt, B. D. L. and McCartney, H. A., 1986. Spore dispersal in relation to epidemic models. In: Plant disease epidemiology. Vol. 1: Population dynamics and management. Eds Leonard, K. J. & Fry, W. E. New York: Macmillan Publishing Company.

36. Gause, G. F., 1934. The struggle for existence. Baltimore: Williams & Wilkins.

37. Gibson, G. J., 1997. Investigating mechanisms of spatio–temporal epidemic spread using stochastic models. *Phytopathol.* **87**, 139.

38. Gibson, G. J. and Austin, E. J., 1996. Fitting and testing stochastic models with application in plant epidemiology. *Plant Pathol.* **45**, 172.

39. Gibson, G. J., Gilligan, C. A., Kleczkowski, A., 1999. Predicting variability in biological control of a plant–pathogen system using stochastic models. *Proc. R. Soc. Lond. B* **266**, 1743.

40. Gilligan, C. A., 1985. Probability models for host infection by soilborne fungi. *Phytopathol.* **75**, 61.

41. Gregory, P. H., 1973. The microbiology of the atmosphere. 2nd ed. London: Leonard Hill.

42. Grimmet, G. R. and Stirzaker, D. R., 1992. Probability and random processes. Oxford Science Publications.

43. Gubbins, S. and Gilligan, C. A., 1997 a. A test of heterogeneous mixing as a mechanism for ecological persistence in a disturbed environment. *Proc. R. Soc. Lond. B* **264**, 227.

44. Gubbins, S. and Gilligan, C. A., 1997 b. Biological control in a disturbed environment. *Phil. Trans. R. Soc. Lond. B* **352**, 1935.

45. Hall, G. and Watt, J. M., 1976. Modern numerical methods for ordinary differential equations. Oxford: Clarendon Press.

46. Hansen, J. P. and McDonald, I. R., 1979. Theory of simple liquids. London: Academic Press.

47. Harris, T. E., 1974. Contact interactions on a lattice. *Ann. Prob.* **2**, 969.

48. Hengeveld, R., 1989. Dynamics of biological invasions. London: Chapman and Hall.

49. Hethcote, H. W. and van den Driessche, P., 1991. Some epidemiological models with nonlinear incidence. *J. Math. Biol.* **29**, 271.

50. Hickman, C. J., 1940. The red core root disease of the strawberry caused by *Phytophthora Fragariae* n.sp. *Journal of Pomology and Horticultural Science* **18**, 89.

51. Hiebeler, D., 1997. Stochastic spatial models: from simulations to mean field and local structure approximations. *J. Theor. Biol.* **187**, 307.

52. Huffaker, C. B., 1958. Experimental studies on predation: dispersion factors and predator–prey interactions. *Hilgardia* **27**, 343.

53. Hughes, G., McRoberts, N., Madden, L. V., Nelson, S. C., 1997. Validating mathematical models of plant–disease progress in space and time. *IMA J. Math. Appl. Med.* **14**, 85.

54. Jeger, M. J., 1986. The potential of analytic compared with simulation approaches to modeling in plant disease epidemiology. In: Plant disease epidemiology. Vol. 1: Population dynamics and management. Eds Leonard, K. J. & Fry, W. E. New York: Macmillan Publishing Company.

55. Jeger, M. J., 1989. Ed. Spatial components of plant disease epidemics. Englewood Cliffs: Prentice Hall.

56. Kikuchi, R., 1951. Theory of cooperative phenomena. *Phys. Rev.* **81**, 988.

57. Kimura, M. and Weiss, G. H., 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**, 561.

58. Kirkwood, J. G., 1935. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **3**, 300.

59. Kleczkowski, A., Bailey, D. J., Gilligan, C. A., 1996. Dynamically–generated variability in plant–pathogen systems with biological control. *Proc. R. Soc. Lond. B* **263**, 777.

60. Konno, N., 1994. Phase transitions of interacting particle systems. Singapore; London: World Scientific.

61. Kramers, H. A. and Wannier, G. H., 1941. Statistics of the two–dimensional ferromagnet. *Phys. Rev.* **60**, 252.

62. Leonard, K. J. & Fry, W. E., 1986. Eds. Plant disease epidemiology. Vol. 1: Population dynamics and management. New York: Macmillan Publishing Company.

63. Levin, S. A. and Durrett, R., 1997. From individuals to epidemics. *Phil. Trans. R. Soc. Lond. B* **351**, 1615.

64. Lewis, M. A., 1997. In: Spatial ecology: the role of space in population dynamics and interspecific interactions. Eds Tilman, D. & Kareiva, P. M. Princeton, NJ: Princeton University Press.

65. Liggett, T. M., 1985. Interacting particles. New York: Springer–Verlag.

66. Liggett, T. M., 1999. Stochastic interacting systems: contact, voter and exclusion processes. Berlin; London: Springer–Verlag.

67. Liu, W., Hethcote, H. W., Levin S. A., 1987. Dynamical behaviour of epidemiological models with nonlinear incidence rates. *J. Math. Biol.* **25**, 359.

68. Liu, W., Levin S. A., Iwasa, Y., 1986. Influence of nonlinear incidence rates upon the behaviour of SIRS epidemiological models. *J. Math. Biol.* **23**, 187.

69. Lotka, A. J., 1925. Elements of physical biology. Baltimore: Williams and Wilkins.

70. Matsuda, H., Ogita, N., Sasaki, A., Sato, K., 1992. Statistical mechanics of populations – the lattice Lotka–Volterra model. *Prog. Theor. Phys.* **88**, 1035.

71. Maule, M. M. and Filipe, J. A. N., 2000 a. Modelling plant epidemics with anisotropic spread: from 2 to 1 dimension. In preparation.

72. Maule, M. M. and Filipe, J. A. N., 2000 b. Relating parameterised heterogeneous mixing to spatial processes in plant epidemics. In preparation.

73. McNeil, D. R., 1972. On the simple stochastic epidemic. *Biometrika* **59**, 494.

74. Minogue, K. P., 1986. Disease gradients and the spread of disease. In: Plant disease epidemiology. Vol. 1: Population dynamics and management. Eds Leonard, K. J. & Fry, W. E. New York: Macmillan Publishing Company.

75. Minogue, K. P., 1989. Diffusion and spatial probability models for disease spread. In: Spatial components of plant disease epidemics. Ed. Jeger, M. J. Englewood Cliffs: Prentice Hall.

76. Minogue, K. P. and Fry, W. E., 1983. Models for the spread of disease: Model description. *Phytopathol.* **33**, 529.

77. Mollison, D., 1977. Spatial contact models for ecological and epidemic spread. *J. R. Statis. Soc. B* **39**, 283.

78. Mollison, D., 1995. Ed. Epidemic models and their relation to data. Cambridge University Press.

79. Mollison, D., Isham, V., Grenfell, B., 1994. Epidemics: models and data. *J. R. Statist. Soc. A* **157**, 115.

80. Morgan, B. J. T., 1984. Elements of simulation. Cambridge University Press.

81. Mundt, C. C. and Leonard, K. J., 1985. A modification of Gregory model for describing plant–disease gradients. *Phytopathol.* **75**, 930.

82. Murray, J. D., 1989. Mathematical biology. Berlin; London: Springer–Verlag.

83. Numerical Algorithms Group, 1993. Library Manual Mark 19. NAG Central Office: Oxford.

84. Okubo, A., 1980. Diffusion and ecological problems: mathematical models. Biomathematics, Vol. 10. Berlin; New York: Springer–Verlag.

85. Pascual, M. and Levin, S. A., 1999. Spatial scaling in a benthic population model with density–dependent disturbance. *Theor. Pop. Biol.* **56**, 106.

86. Renshaw, E., 1991. Modelling biological populations in space and time. Cambridge University Press.

87. Reynolds, K. M. and Madden, L. V., 1988. Analysis of epidemics using spatio–temporal autocorrelation. *Phytopathol.* **78**, 240.

88. Ripley, B. D., 1987. Stochastic simulation. New York; Chichester: Wiley.

89. Russell, G., 1996. Spatial aspects of modelling in applied biology. In: Modelling in applied biology: spatial aspects. Aspects of Applied Biology 46, 1.

90. Sato, K., Matsuda, H., Sasaki, A., 1994. Pathogen invasion and host extinction in lattice structured populations. *J. Math. Biol.* **32**, 251.

91. Severo, N. C., 1969. Generalizations of some stochastic epidemic models. *Math. Biosci.* **4**, 395.

92. Shaw, M., 1994. Modeling stochastic processes in plant pathology. *Annu. Rev. Phytopathol.* **32**, 523.

93. Shaw, M., 1995. Simulation of population expansion and spatial pattern when individual dispersal distributions do not decline exponentially with distance. *Proc. R. Soc. Lond. B* **259**, 243.

94. Stanley, E. H., 1972. Introduction to phase transitions and critical phenomena. Oxford University Press.

95. Van de Lande, H. L. and Zadocks, J. C., 1999. Spatial patterns of spear rot in oil palm plantations in Surinam. *Plant Pathol.* **48**, 189.

96. Vanderplank, J. E., 1963. Plant diseases: epidemics and control. New York: Academic Press.

97. Volterra, V., 1926. Fluctuations in the abundance of a species considered mathematically. *Nature* **118**, 558.

98. Wolfram, S., 1986. Theory and application of Cellular Automata. Singapore: World Scientific.

99. Xu, X.-M. and Ridout, M. S., 1996. Analysis of disease incidence data using a stochastic spatial–temporal simulation model. In: Modelling in applied biology: spatial aspects. Aspects of Applied Biology 46, 155.

100. Xu, X.-M. and Ridout, M. S., 1998. Effects of initial epidemic conditions, sporulation rate and dispersal gradient on the spatio–temporal dynamics of plant disease epidemics. *Phytopathol.* **88**, 1000.

101. Zadocks, J. C. and van den Bosch, F., 1984. On the spread of plant disease: a theory on foci. *Annu. Rev. Phytopathol.* **32**, 503.

102. Zawolek, M. W., 1993. Shaping a focus: wind and stochasticity. *Neth. J. Pl. Path.* **99**, 241.