Chapter Seven

Spatial Models

It is intuitive that in most circumstances disease transmission is predominantly a localized process. For directly transmitted diseases, for example, transmission is most likely between individuals with the most intense interaction, which generally implies those in the same location. Additionally, movement of individuals between population centers facilitates the geographical spread of infectious diseases. This chapter is concerned with capturing these host population characteristics, enabling us to address issues such as: determining the rate of spatial spread of a pathogen, calculating the influence of large populations on smaller ones, and finding optimally targeted control measures that take into account the local nature of spatial transmission. Generally, models of this sort operate by partitioning the population according to the spatial position of hosts, such that nearby hosts are grouped together and interact more strongly. A wide variety of model formats have been developed to accomplish this, with the primary differences being the scale at which hosts are aggregated. Section 7.7 outlines when each of the model types can be used, although no definitive rules exist. Rigorous analytical results for spatial epidemiological models remain rare. Since the late 1980s, however, the increasing ease of access to computational power has permitted the detailed simulation of such models (Levin et al. 1997). Frequently, these models incorporate stochasticity, so readers may wish to familiarize themselves with Chapter 6 before continuing.

We will introduce the main types of spatial models in some detail, outline the general methodology, comment on their strengths and weaknesses, and consider how such models have been used to explore the spatial spread of real-world epidemics. The following three motivating examples will illustrate how the behavior of the infectious disease and the host, and the amount of knowledge available, dictates the form of model.

The 2001 foot-and-mouth epidemic in the United Kingdom was a clear example of how space can play a significant role in disease dynamics. Cases were predominantly restricted to three regions: Cumbria, Devon, and the Welsh borders, with most transmission occuring within 3 kilometers from the source farm. Given that the location of all farms in the United Kingdom was known, and that accurate predictions (rather than generic insights) were required, two of the models used during the epidemic (Keeling et al. 2001b; Morris et al. 2001) were individual-based and explicitly spatial, using the known location of farms and the estimated rate of transmission as a function of distance to simulate the epidemic.

If we are interested in the spread of human disease, such as pandemic influenza (Grais et al. 2003) or SARS (Riley et al. 2003), then models need to focus on the dynamics at a large geographical scale. Given our limited knowledge about human movement patterns and interactions, it is typically impractical to simulate each individual within a population, although this has recently been attempted for both small (Halloran et al.

2002; Longini et al. 2005) and large population sizes (Ferguson et al. 2005, Ferguson et al. 2006). One practical limitation of these models is the difficulty with which we can assess the sensitivity of their predictions to perturbations in the social network structure. Instead, it is often plausible to assume random mixing within a localized community and reduced mixing between the communities—a so-called metapopulation model. We may, for example, wish to subdivide the population by town/city, by county, or by state, depending upon our level of knowledge and the detail of results required (see, for example, Viboud et al. 2006).

Finally, for diseases of wildlife or plants, such as tuberculosis (TB) in badgers (Shirley et al. 2003), rabies in foxes (Murray et al. 1986) or raccoons (Smith et al. 2002), Dutch elm disease (Swinton and Gilligan 1996), or Sudden Oak Death (Kelly and Meentemeyer 2002), there may not be a natural partitioning of the host population. Instead, it is frequently assumed that individuals are either uniformly or randomly distributed, with their density reflecting landscape and environmental factors. In such cases, continuous-space models, phrased as partial differential equations (PDEs) or integro differential equations (IDEs), can be used (Kot 2001). Again, it would be unfeasible to model every badger sett within the United Kingdom or every raccoon in the United States, and naive to assume that either creature respects county or state boundaries. However, due to limitations arising from the spatial resolution of empirical data, which may be aggregated at the county or state level, it may be necessary to model the host population at a similarly coarse scale (Smith et al. 2002).

The type of model used is directly dependent on the host organism, our degree of knowledge about its behavior, and the scale we wish to consider.

7.1. CONCEPTS

A variety of models can be used to study the spatial spread of pathogens, and although each has its own specific aspects, a range of concepts are shared. We first discuss these elements, so that the similarities and differences between the models will be more apparent, and to introduce the language of spatial processes.

7.1.1. Heterogeneity

Spatial heterogeneity refers to differences between populations or individuals at different geographical locations. Such heterogeneities can arise from two sources. Underlying (environmental) heterogeneities describe spatial differences in the fundamental forces governing the population dynamics. For example, wildlife populations in different locations may experience differing habitat conditions that may affect demographic rates, or different human populations may have different social structures leading to variation in disease transmission rates (Finkenstädt and Grenfell 1998; Grenfell and Bolker 1998; van Buskirk and Ostfeld 1998; Auvert 2000). Such underlying heterogeneities are common in the real world, but are frequently ignored in models due to the extra complexity they introduce and due to a lack of available data. If quantitatively precise predictions are required from models, however, it is often vitally important that such underlying heterogeneities are considered (Keeling et al. 2001b; Smith et al. 2002). The second form

of heterogeneity is emergent and describes observed differences in population structures arising from dynamical processes, such as stochasticity, or differences in movement between populations (Hassell et al. 1991; Rhodes and Anderson 1996; Green and Sadedin 2005). In general, this second form of heterogeneity is greatest between populations that experience large amounts of stochasticity, have very different underlying parameters, and have little transfer of infection between them.

Heterogeneity can describe either the underlying differences between two populations, or the emerging dynamic differences in the population levels (such as the proportion of the population that are infectious).

A convenient measure of observed heterogeneity is provided by estimating *correlations*—they quantify the degree to which the dynamics in two (or more) populations behave in the same manner. Simply put, correlations help to establish whether epidemics in different populations are synchronized or out of phase (Grenfell and Bolker 1998; Rohani et al. 1999; Grenfell et al. 2001). If we let I_i denote the time series documenting the prevalence of an infection in population i, then the correlation between epidemics in two populations is calculated as:

$$C_{12} = \frac{(I_1(t) - \overline{I_1})(I_2(t) - \overline{I_2})}{\sqrt{\text{var}(I_1)\text{var}(I_2)}}. \tag{7.1}$$

Here, $\overline{I_i}$ refers to the mean infection prevalence (averaged over time) in population i. If the fluctuations in prevalence over time in the two populations are either identical or directly proportional ($I_1 \propto I_2$), then the correlation attains its maximum value of 1. If epidemics in the two populations are independent, then the correlation is zero. If the outbreaks are out of phase, then C_{12} is negative. Given time-series data on the number of cases in two populations, we are predominately interested in the average correlation over a given period, rather than the instantaneous value that is subject to short-term stochastic fluctuations. The correlation cannot be defined for deterministic populations at their equilibrium values (because both population levels are constant and the variance is zero), and therefore, in general, correlations are usually associated with stochastic or seasonally forced systems.

Correlations provide a quantitative measure of the differences between populations: A positive/negative correlation indicates that epidemics are spatially synchronous/asynchronous.

Although the standard correlation (equation (7.1)) measures the heterogeneity generally derived from the stochastic nature of the epidemic process, heterogeneities can also arise due to *traveling-waves*. Consider the spread of West Nile virus across the United States from New York in 1999 to the West Coast in 2003 (see Chapter 4 for a more detailed description of West Nile virus). The observed heterogeneities in incidence on the East and West Coast are unlikely to be a result of either inherent habitat differences, or due to stochasticity. They simply reflect the fact that the disease emerged in the east and travelled west. To quantify this traveling-wave type of heterogeneity we need

to use *lagged correlations*:

$$C_{12}^{\tau} = \frac{(I_1(t+\tau) - \overline{I_1})(I_2(t) - \overline{I_2})}{\sqrt{\text{var}(I_1)\text{var}(I_2)}}. \tag{7.2}$$

If a traveling wave is observed, then the value of τ that maximizes the lagged correlation, C_{12}^{τ}, should increase with the separation, d, between two populations. If the traveling wave moves with constant velocity, c, then $\tau_{max} = d/c$, which can be derived from the fact that the time taken for the wave to travel a distance, d, is the distance divided by the velocity. More advanced versions of this approach have been successfully used by Grenfell et al. (2001) to identify traveling waves of measles infection in England and Wales spreading from large population centers like London to the surrounding smaller communities.

7.1.2. Interaction

Consider the behavior of an infectious disease within several human populations. If there is no interaction (movement) between the populations, then their dynamics will be independent and hence the correlation between them will be zero (assuming no other synchronizing mechanisms, such as seasonal forcing or climatic factors). However, movement of hosts between populations, with the associated risk of disease transmission, can couple dynamics. The way in which we choose to model this interaction should reflect the behavior of the host and the scale at which our model operates.

One of the simplest means of modeling the interaction between (for example) two populations is for susceptible individuals at one location to experience an additional force of infection due to infectious individuals at the other. This would represent a phenomenological approach to spatial modeling and we frequently refer to the strength of such interactions as the level of *coupling* between the populations. The greater the coupling, the more each population is impacted by the transmission dynamics of the other and the higher the level of correlation and synchrony.

Interaction or coupling between different spatial locations allows infection to spread and acts to synchronize the epidemic dynamics at the two locations.

It is intuitive that in many situations the interaction between two populations should decrease with the distance, d, between them. This type of behavior can be captured by introducing a *transmission kernel*, K, which modifies the coupling term and is a function of the distance between two populations. Common examples of transmission kernels include exponential ($K \propto \exp(-Ad)$), Gaussian ($K \propto \exp(-Ad^2)$), or power-law ($K \propto d^{-A}$) (Erlander and Stewart 1990; Gibson 1997a,b; Keeling et al. 2004b; Xia et al. 2004), with the precise form chosen determined by the observed dynamics. Estimating the kernel form and parameters is a very difficult but important problem; although the more common, short-distance transmission events determine the basic reproductive ratio, it is the long-distance tail of the kernel that determines the eventual speed of a traveling wave of invading pathogen (Diekmann 1978; van den Bosch et al. 1990; Mollison 1991; Shaw 1995; Lewis 2000; Xia et al. 2004). However, at long distances the kernel is usually very small, so there will be only limited amounts of data for the estimation processes. These rare jumps to new areas cannot be ignored, however, because they are often vitally important to the invasion process.

The reduction in transmission risk with distance is captured by a transmission kernel, which is frequently assumed to be either exponential, Gaussian, or power-law.

7.1.3. Isolation

Isolation is another factor that is common in a wide range of models and real scenarios. It simply refers to the situation when a group of hosts is protected from the risk of transmission due to their spatial separation from an infectious source. For example, we might consider communities that have few contacts with the outside world as being isolated from the general pool of infectious individuals. Alternatively, animal populations that are separated by large distances can be epidemiologically isolated from other populations. The existence of isolated populations can have a profound impact on parameterization; if isolated populations are included in an estimate of disease parameters, their rarity of transmission will bias the results.

7.1.4. Localized Extinction

In any stochastic population model, there is always the risk that the disease will, by chance, become extinct and this risk increases as the host population size gets smaller (see Chapter 6). This is where the spatial resolution of study becomes important. For example, if we consider the aggregate epidemics of an infectious disease like measles in the prevaccine era in the whole of England and Wales, then the probability of witnessing a fade out is almost zero. However, as we examine case reports in increasingly smaller cities and towns, the frequency of local extinctions increases, with the smallest population centers experiencing fade outs in between epidemics sparked by the introduction of infection from cities where it is endemic. The likelihood of such "recolonization events" is influenced by the synchrony of measles epidemics and the coupling between subpopulations (Bolker and Grenfell 1996; Earn et al. 1998; Rohani et al. 1999; Keeling 2000b; Hagenaars et al. 2004). As a result, overall population persistence is determined by a key relationship between the subpopulation size, the degree of interaction, and the strength of asynchrony. The precise details of this relationship remain largely unclear, but some of the complexities are discussed below when metapopulations are described (Section 7.2.3).

Due to the smaller subpopulation sizes often involved in spatial models, localized extinctions are common. Large-scale eradication is prevented by coupling between subpopulations leading to the reintroduction of infection into disease-free areas.

7.1.5. Scale

Two forms of scale are important for spatial models: (1) the scale of interaction, and the (2) the scale of simulation.

The majority of spatial models make some assumption about the spatial scale of interaction and the scale at which the population can be subdivided. Although there is rarely a "correct" scale, it is clear that using too fine a scale and hence creating many subpopulations can be computationally prohibitive, whereas aggregating at too large a scale

can eliminate the spatial effects that are of primary interest. For individual-based models (which operate at the level of the individual) and metapopulation models (which consider communities as the aggregate unit), the scale is generally fixed. However, for both lattice-based and continuous-space models (PDEs and IDEs), the question of scale is much more subtle. In such models, it is frequently assumed that each population interacts only with a limited number of other populations (usually within a prescribed distance), and this set of populations is referred to as the (interaction) *neighborhood*. In general, the use of a finite neighborhood is an approximation to the true dynamics, but may greatly increase the speed of any spatial simulation.

When dealing with a particular "real-world" problem, such as foot-and-mouth in the United Kingdom or SARS in North America, the spatial extent of our model is fixed by the problem. However, in more abstract or generic situations, the scale at which the spatial model operates is less constrained. Intuitively, we wish to model at a sufficiently large scale so that the full range of dynamics is observed and the scale of the model has limited impact on our results. However, models that are too "big" may be slow to compute, thereby limiting their usefulness. Finding the correct scale usually involves assessing whether the salient behavior of large-scale models is still captured by models of a smaller, more computationally manageable scale. Therefore, at least initially, large-scale simulations are required to determine the unconstrained dynamics. Although a number of dynamical systems techniques exist to find the optimal, most informative scale (Mead 1974; Keeling et al. 1997a; Pascual et al. 2001), the answer is often context dependant.

Assessing the fine scale at which individuals are aggregated and the larger scale at which simulations are performed should either be based on sound epidemiological knowledge or achieved by comparing simulated results across a range of scales.

7.2. METAPOPULATIONS

Metapopulations are one of the simplest spatial models, but are also one of the most applicable to modeling many human diseases. The metapopulation concept is to subdivide the entire population into distinct "subpopulations", each of which has independent dynamics, together with limited interaction between the subpopulations. This approach has been used to great effect within the ecological literature; a comprehensive guide to the broader applications of this approach is given by Hanski and Gilpin (1991, 1997) and Hanski and Goggiotti (2004). For disease-based metapopulation models, a suitable modified version of the SIR equation (Chapter 2) would be:

$$\begin{aligned} \frac{dX_i}{dt} &= \nu_i N_i - \lambda_i X_i - \mu_i X_i, \\ \frac{dY_i}{dt} &= \lambda_i X_i - \gamma_i Y_i - \mu_i Y_i, \end{aligned} \tag{7.3}$$

where the subscript i defines parameters and variables that are particular to subpopulation i. The force of infection, λ_i incorporates transmission from both the number of infecteds within subpopulation i and the coupling to other subpopulations. In this general formulation, the demographic and epidemiological parameters may vary between subpopulations,

reflecting differences in the local environments (Finkenstädt and Grenfell 1998; Grenfell and Bolker 1998; Langlois et al. 2001; Broadfoot et al. 2001).

Metapopulations provide a powerful framework for modeling disease dynamics for hosts that can be naturally partitioned into spatial sub-units.

The precise relationship between the force of infection for population i and the number of infectious individuals in population j depends on the assumed mechanism of transmission and the strength of interaction between the two populations. In general terms, the force of infection can be written as a sum:

$$\lambda_i = \beta_i \sum_j \rho_{ij} \frac{X_j}{N_i},$$

where the coefficients, ρ, are a measure of the strength of interaction between populations. Specifically, ρ_{ij} measures the relative strength of transmission *to* subpopulation i *from* subpopulation j. An important aspect of this formulation concerns the precise scaling with population size in the expression of λ_i. The equation above contains N_i in the denominator, which reflects the implicit assumption that transmission takes place in population i, presumably resulting from the movement of an infectious individual from population j. Alternatively, the assumption that transmission is due to a susceptible individual from population i picking up the infection during a temporary visit to population j would be incorporated by placing N_j in the denominator. The assumptions implicit in the coupling interaction are discussed more fully in Section 7.2.1.

The force of infection within a subpopulation can be expressed as a weighted sum of the prevalence in all populations.

We now explore the differences between deterministic and stochastic versions of this metapopulation model. Consider two large, fully susceptible populations ($S_1 = S_2 = 1$), with $\rho_{ii} = 1$ and ρ_{ij} much less than 1. (We will assume that the two populations are the same size, which simplifies the form of the coupling interaction, and ignore the effects of demography.) We start with an infectious disease solely in population 1, which exhibits a standard epidemic curve (Chapter 2), because the coupling between populations is assumed to be small. In the deterministic framework, ignoring births and deaths, the early dynamics of population 2 (before the proportion of susceptibles drops significantly, $S_2 \approx 1$) will be given by:

$$\frac{dI_2}{dt} = \beta_2 \rho_{21} I_1 + \beta_2 I_2 - \gamma_2 I_2.$$

This equation can be solved using the "integrating factor" (see, for example, Strang 1986) to obtain an expression for I_2 through time

$$I_2(t) = \int_0^t \beta_2 \rho_{21} I_1(s) \exp([\beta_2 - \gamma_2]s) ds. \tag{7.4}$$

This expression represents the exponential growth of prevalence in population 2 from time 0 to time t, due to the presence of $I_1(s)$ infecteds in population 1 at time s. The deterministic equation (7.4) has two main implications: the disease is "present" in population 2 from the start of the epidemic (in population 1), and the early infinitesimal infections that arrive in population 2 trigger an exponential growth at rate $\beta_2 - \gamma_2$.

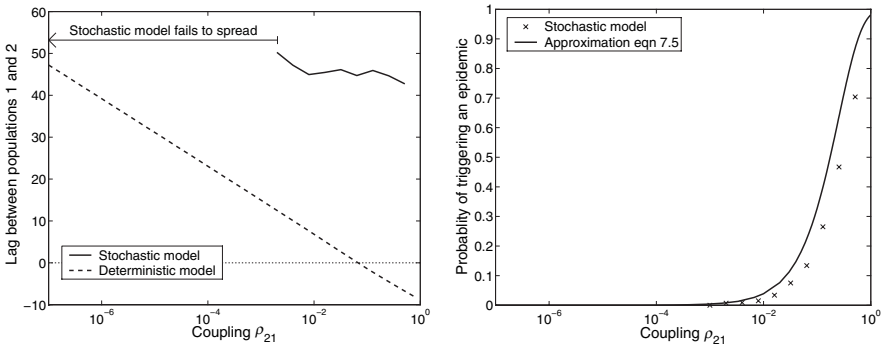


Figure 7.1. For a metapopulation with just two populations we examine the effect of coupling ρ_{21} on the dynamics of the epidemic in population 2. Population 1 is modeled deterministically and is initialized with $I(0) = 10^{-5}$, $S(0) = 1 - I(0)$. The left-hand graph shows the delay between the peak of the epidemic in population 1 and the peak in population 2, where population 2 is modeled either stochastically (solid line) or deterministically (dashed line) and is initially disease-free. The right-hand graph is the probability that a major epidemic is triggered in the stochastic population (crosses) compared to the analytical approximation, equation (7.5). ($\mu = \nu = 0$, $1/\gamma = 14$ days, $\beta = 0.3571$ per day $\Rightarrow R_0 = 5$, $\rho_{ii} = 1$, $\rho_{12} = 0$. Stochastic results are the average of 1,000 realizations, with population 1 treated deterministically, $N_2 = 10^5$).

In the analogous stochastic formulation, model behavior is significantly different. The probability that a major epidemic is triggered in population 2 is given by:

$$\begin{aligned}
\mathbb{P}(\text{epidemic}) &= \sum_{n=1}^{N} \mathbb{P}\,(\text{cases in 1 cause } n \text{ cases in subpopulation 2}) \times \\
&\qquad \mathbb{P}\,(n \text{ initial cases lead to a major epidemic}), \\
&= \sum_{n=1}^{N} \exp\left(-\beta_2\rho_{21}\int_0^\infty I_1(s)ds\right) \frac{\left(\beta_2\rho_{21}\int_0^\infty I_1(s)ds\right)^n}{n!} \times \left[1-\left(\frac{\gamma}{\beta_2}\right)^n\right], \\
&= 1-\exp\left(-\beta_2\rho_{21}\left[1-\frac{\gamma}{\beta_2}\right]\int_0^\infty I_1(s)ds\right) \qquad (7.5) \\
&< 1-\exp\left(-\beta_2\rho_{21}/\gamma\right).
\end{aligned}$$

So, in the stochastic formulation, if the coupling ρ_{21} between populations is small enough, there is a good chance that the epidemic will fail to spread (Park et al. 2002). When ρ_{21} is larger, although the pathogen may eventually spread, there may still be a significant delay before other populations are exposed. Hence, in stochastic metapopulation models, the spread of infectious disease is slower than for the deterministic counterpart.

This principle is illustrated in Figure 7.1. The pathogen is introduced in population 1 only, with $\rho_{ii} = 1$, and we measure the lag between the peak of the epidemic in population 1 and the subsequent peak in population 2. In both models, the delay between the peaks

decreases as the coupling ρ_{21} increases. For these parameters, however, it is not until the coupling exceeds approximately 0.01 that an epidemic has a significant chance of being triggered in the stochastic model. Two interesting aspects should be noted. The first is that for high levels of coupling, the lag in the deterministic model is negative (population 2 peaks first); this is because the force of infection in population 2 is boosted by the coupling from population 1, giving it a faster rate of epidemic growth (note that we have assumed that $\rho_{12} = 0$, so population 1 cannot receive infection from population 2). Second, the expected average lag predicted by stochastic simulations is far longer than the lag from the deterministic equations due to the chance nature of the initial transmission between populations.

With stochastic metapopulations, the spread of pathogen between subpopulations is reduced compared to the equivalent deterministic model.

7.2.1. Types of Interaction

We now turn our attention to the types of spatial transmission that are associated with different hosts and different types of movement. Although for convenience we label these as plant, animal, and human-commuter interactions, these descriptions are not rigid but rather illustrate the underlying host dynamics that generate the transmission terms. Hence, animal subpopulations that actually interact via occasional encounters would be better modeled as commuters, whereas animal subpopulations where the interaction is due to the wind-borne pathogen spread are more realistically modeled using the plant formulation.

In some circumstances, the interaction between subpopulations can be parameterized from direct observation, as is the case for commuter movements between communities (see Section 7.2.1.3) or from comparison to the recorded epidemic patterns (Wallace and Wallace 1993; Smith et al. 2002; Xia et al. 2004). Alternatively, the interaction terms between subpopulations are often assumed to obey a simple distance relationship (Erlander and Stewart 1990; Finkenstädt and Grenfell 1998).

7.2.1.1. Plants

The most obvious defining feature of plants (from an epidemiological perspective) that separates them from other hosts is that they do not move. This means that any spatial transmission must be wind- or vector-borne. We therefore retain the formulation

$$\lambda_i = \beta_i \sum_j \rho_{ij} I_j \tag{7.6}$$

and consider the coupling, ρ, as a function that decreases with the distance between the subpopulations, and for simplicity set $\rho_{ii} = 1$ (Park et al. 2001, 2002; Thrall et al. 2003). We can now calculate R_0 for infectious individuals in population i, as the expected number of secondary cases generated in all subpopulations:

$$R_0^i = \sum_j \frac{\beta_j \rho_{ji}}{\gamma_i}. \tag{7.7}$$

Note that the coupling term is now ρ_{ji} because we are concerned with transmission to j from i. It is important to realize that with this model, the addition of extra subpopulations

(with extra hosts) increases R_0. Intuitively, this is because more external populations can "capture" wind-borne pathogen particles that otherwise would not have contributed to the transmission process. However, dividing one subpopulation into two should not have the same effect—if population j is divided forming two new populations k and l, then $\rho_{ji} = \rho_{ki} + \rho_{li}$ so that R_0 remains constant. It is only the inclusion of additional host populations that can raise R_0.

The concept of a metapopulation is one that has been readily used by the plant research community, such that a variety of data exist on the distribution of diseases in plants (Burdon et al. 1995; Ericson et al. 1999) with a strong focus on the genetic specialization of the pathogen to local populations (Burdon and Thrall 1999; Bergelson et al. 2001).

Such coupling mechanisms are not exclusively for plant hosts, but can be applied to any sessile population with wind- or vector-borne pathogens. Hence, this form of metapopulation model is ideal for describing the spatial dynamics of livestock diseases, where each farm is a subpopulation and transmission between farms can either be wind-borne or due to the movement of people, vehicles, or animals (Keeling et al. 2001b; Ferguson et al. 2001a; Section 7.5.2).

For plants and other sessile hosts, coupling generally decreases with distance, mimicking the effects of wind- or vector-dispersal. Adding an extra subpopulation generally increases R_0 because more pathogens can be intercepted by the additional hosts.

7.2.1.2. Animals

For many animal populations, it is plausible to assume that the spread of disease is due to the migration or permanent movement of individuals. The simplest means of modeling this is to allow animals to randomly move between subpopulations (Foley et al. 1999; Broadfoot et al. 2001; Fulford et al. 2002), although other assumptions based on known dispersal behavior of specific species leading to different spatio-temporal dynamics may be more appropriate (Gudelj et al. 2004). The metapopulation SIR-type model is then:

$$\begin{aligned} \frac{dX_i}{dt} &= \nu_i - \beta_i X_i Y_i - \mu_i X_i + \textstyle\sum_j m_{ij} X_j - \sum_j m_{ji} X_i, \\ \frac{dY_i}{dt} &= \beta_i X_i Y_i - \gamma_i Y_i - \mu_i Y_i + \textstyle\sum_j m_{ij} Y_j - \sum_j m_{ji} Y_i. \end{aligned} \tag{7.8}$$

This is online program 7.1

Here, coupling is governed by the parameter m_{ij}, which measures the rate at which hosts migrate to subpopulation i from j—and therefore captures both emigration and immigration. We have assumed density-dependent transmission in equation (7.8), reflecting the common assumption about wildlife diseases—although frequency-dependent transmission could easily be accommodated. It is frequently assumed that the movement rates balance, $m_{ij} = m_{ji}$, so that the subpopulation sizes are maintained—there is no reason why this has to be the case and some subpopulations could act as sources of animals that colonize less favorable habitats. In this coupling framework (and assuming that β and γ are population independent and population sizes are equal or transmission is frequency dependent), the basic reproductive ratio is $\frac{\beta N}{\gamma+\mu}$ and is independent of the coupling strength. Intuitively, this is because each infectious animal transmits at a constant rate irrespective of which population it is in, and so always generates the same average number of secondary cases.

Compared to the plant models above, the movement of animals has two very different components: (1), it is the direct movement of infected animals that spreads the pathogen, and (2) the movement of susceptibles can help prevent stochastic extinctions (see Section 6.3.3) in heavily infected subpopulations. In this way, the coupling implied by animal movements is more variable than that due to wind or vector transmission assumed for plant infections. For the plant model (7.6), there is a continual force of infection from subpopulation 1 to subpopulation 2 (and vice versa); however, for the animal-based model (7.8) transmission between subpopulations occurs only following the movement of an animal. Therefore, in a stochastic framework either no infected animals have moved recently, in which case there is no transmission between populations, or an infected animal has moved, in which case the transmission is reasonably strong.

Models of animal diseases usually capture the transmission of infection by the permanent immigration and emigration of hosts. In these models R_0 is generally independent of the coupling because each host transmits infection at a constant rate.

7.2.1.3. Humans

For human populations, permanent relocation from one population to another is sufficiently rare that it may be ignored as an epidemiologically significant force. Instead, it is more natural to think about commuters spreading the disease (Keeling and Rohani 2002). Commuters live in one subpopulation but travel occasionally to another subpopulation. We therefore label X_{ij}, Y_{ij}, and N_{ij} as the number of susceptibles, infecteds, and total hosts currently in population i that live in population j. When there are multiple communities within the metapopulation, and when the populations are of different sizes or the strengths of interaction differ, it is more informative to return to first principles to calculate the dynamics. From the standard SIR models (Chapter 2) we consider the number of individuals of each type (S, I, and R) in each spatial class:

This is online program 7.2

$$
\begin{aligned}
\frac{dX_{ii}}{dt} &= \nu_{ii} - \beta_i X_{ii} \frac{\sum_j Y_{ij}}{\sum_j N_{ij}} - \sum_j l_{ji} X_{ii} + \sum_j r_{ji} X_{ji} - \mu_{ii} X_{ii}, \\
\frac{dX_{ij}}{dt} &= \nu_{ij} - \beta_i X_{ij} \frac{\sum_j Y_{ij}}{\sum_j N_{ij}} + l_{ij} X_{jj} - r_{ij} X_{ij} - \mu_{ij} X_{ij}, \\
\frac{dX_{ii}}{dt} &= \beta_i X_{ii} \frac{\sum_j Y_{ij}}{\sum_j N_{ij}} - \gamma Y_{ii} - \sum_j l_{ji} Y_{ii} + \sum_j r_{ji} Y_{ji} - \mu_{ii} Y_{ii}, \\
\frac{dY_{ij}}{dt} &= \beta_i X_{ij} \frac{\sum_j Y_{ij}}{\sum_j N_{ij}} - \gamma Y_{ij} + l_{ij} Y_{jj} - r_{ij} Y_{ij} - \mu_{ij} Y_{ij}, \\
\frac{dN_{ii}}{dt} &= \nu_{ii} - \sum_j l_{ji} N_{ii} + \sum_j r_{ji} N_{ji} - \mu_{ii} N_{ii}, \\
\frac{dN_{ij}}{dt} &= \nu_{ij} + l_{ij} N_{jj} - r_{ij} N_{ij} - \mu_{ij} N_{ij},
\end{aligned}
\tag{7.9}
$$

where l_{ij} measures the rate that individuals leave their home population j and commute to population i, and r_{ij} measures the rate of return. Frequently, in many human disease scenarios the parameters l and r can be found from commuter movement data or travel statistics (Grais et al. 2003; Cliff and Haggett 2004). Other parameters are allowed to

depend on both the home and current location—hence v_{ij} refers to individuals who live in location j but are born in location i. In these equations we have assumed frequency dependent transmission—as is the normal for human diseases—with $\sum_j N_{ij}$ giving the number of individuals currently in population i.

Using the formulation of equation (7.9), Figure 7.2 provides an example of an infectious disease spreading through the 67 counties of Great Britain. The simulations are initialized with the entire population of each county (as recorded in the 1991 census) being susceptible, and 10 infectious cases are placed in Inner London. The rates at which individuals commute l_{ij} are also taken from the 1991 census, while individuals are assumed to return home relatively quickly, at a rate of $r_{ij} = 2$ per day. Although this model is parameterized at a national level in terms of regular commuter movements, a similar modeling framework could be used to deal with the more irregular long-distance travel that can spread infection around the globe (Wilson 2003; Grais et al. 2003; Cliff and Haggett 2004).

Many factors emerge from Figure 7.2 that could be intuitively obvious without resorting to a spatial model: The first county to suffer a major epidemic is the source of initial infection, Inner London. Outer London peaks next due to its tight coupling to Inner London, and Outer London experiences the largest epidemic due to it having the largest population size. However, some elements may be more surprising: Greater Manchester and the West Midlands (along with West Yorkshire and Strathclyde) are among the last counties to have a sizable number of cases—despite their frequent interactions with London. This is attributable to the size of the population in these counties, such that although the infection may arrive relatively early it takes many weeks before the epidemic reaches its peak. Despite the clear delays between epidemics in different counties, within the first 50 days the infection has reached most areas (middle graph), after which time commuting plays a very minor role in the dynamics. This highlights the importance of rapidly imposing movement restrictions if we wish to curtail the spatial spread of infection. Additionally from this graph, it can be seen that the natural time to extinction is more than 300 days despite the fact that prevalence has dropped to very low levels within half this time. Finally, and as expected from simple models (such as shown in Figure 7.1), a deterministic version of this model predicts a faster speed of disease spread.

7.2.1.4. Commuter Approximations

Although equation (7.9) provides a full mechanistic description of the disease behavior, a large number of equations are frequently involved ($3n^2$ equations for n populations), and so it is informative to relate this model to the simpler ones defined earlier. In fact, Keeling and Rohani (2002) showed that for two populations of equal size and equal epidemiological characteristics, equation (7.9) can be simplified by assuming that all commuter movements are very rapid. In this case, the force of infection experienced by population i can be written as:

$$\lambda_i = \beta_i\big((1-\rho)I_i + \rho I_j\big) \qquad j \neq i. \tag{7.10}$$

The coupling parameter, ρ, can be defined in terms of the mechanistic movement of individuals by:

$$\rho = 2q(1-q), \tag{7.11}$$

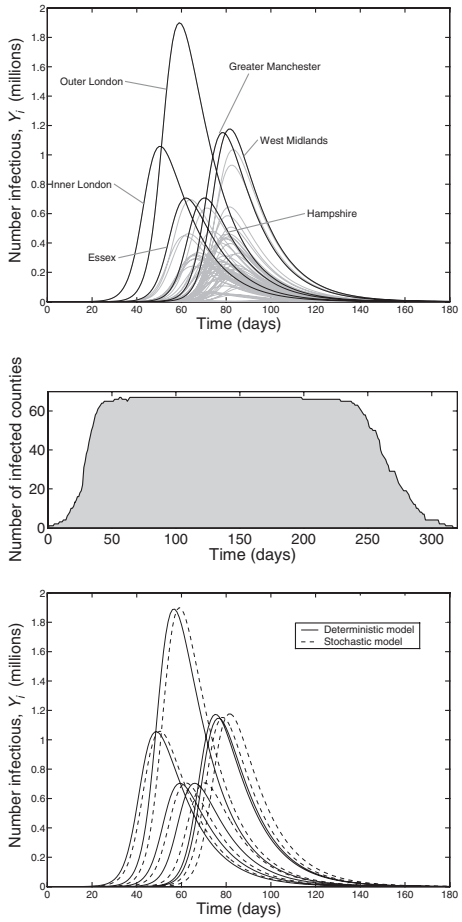


Figure 7.2. Deterministic and stochastic results for an infection spread through the 67 counties of Great Britain. The epidemic is initialized with 10 cases in Inner London, and is spread by commuter movements. The population size and rate of commuting is taken from the 1991 census database, and all trips are considered to be of short duration, $1/r = 0.5$ days. The top figure shows the county-level epidemics from a single stochastic iteration, with six counties highlighted. The middle graph shows the number of counties with infection from the same stochastic model. The bottom graph compares the deterministic solution (solid line) of equations (7.9) with the stochastic model (dashed line) for the six counties highlighted in the top graph. ($\mu = \nu = 0$, $1/\gamma = 14$ days, $\beta = 0.3571$ per day $\Rightarrow R_0 = 5$).

where q is the proportion of the time that individuals spend away in the other population,

$$q = l_{ij}/(r_{ji} + l_{ij}) = l_{ji}/(r_{ij} + l_{ji}).$$

Equation (7.11) is derived from the fact that when *either* the susceptibles of one population or infecteds of the other (but not *both*) move, there is a transfer of pathogen. We therefore find that ρ and hence the transfer of infection is maximized when individuals spend equal amounts of time in both home and away populations, $q = \frac{1}{2}$.

From this commuter approximation (equation (7.10)) it is clear that although increased coupling, ρ, leads to greater transmission between subpopulations, it weakens the transmission within subpopulations, therefore making R_0 independent of the coupling strength. Again, this type of disease transmission does not apply only to human commuters; Swinton et al. (1998) used a similar model to describe the spread of phocine distemper virus through harbour seals in the North Sea. In this context, haul-out beaches, where seals leave the water and congregate, act as natural subpopulations and infection was spread spatially by the occasional visits of seals to nearby beaches—thus the spread of infection is much closer to commuter-type movements than permanent migration usually associated with animals.

The spread of human diseases is best captured by the rapid commuter movements of individuals from their home subpopulation to another subpopulation and back again—requiring us to model both the current location and home location of individuals. When commuter movements are of short duration, this can be approximated by simple coupling. In these models, R_0 is independent of the coupling.

7.2.2. Coupling and Synchrony

Although coupling and the interaction between populations is key for the spatial invasion and spread of a disease, it also affects the endemic dynamics. In particular, the correlation between the disease dynamics in two subpopulations is generally a sigmoidal function of the interaction between them. Figure 7.3 shows the correlation against interaction strength for the four coupling mechanisms discussed above: equations (7.6), (7.8), (7.9), and (7.10). These results echo a general finding within metapopulation models, that the "interesting" spatial dynamics occur when the interaction is between 10^{-3} and 0.1 (Bolker and Grenfell 1995). When the interaction between subpopulations is too small the dynamics are effectively independent and spatial structure is unimportant, whereas when the interaction is too large the dynamics are synchronized, therefore, the subpopulations act like one large well-mixed population and again spatial structure is epidemiologically unimportant. The most notable factor in Figure 7.3 is the similarity between the results for different coupling mechanisms; once correctly scaled the correlation is a function of the interaction strength, with both the form of interaction and population size playing a relatively minor role. The scalings between interaction strengths are approximately:

$$\rho_{\text{plants}} \sim \frac{m}{\gamma + m} \sim 2\frac{l}{r + l}\left(1 - \frac{l}{r + l}\right) \sim \rho_{\text{commuters}}.$$

Hence, whereas commuter movements get scaled by their effective length of stay, permanent migration is multiplied by the average infectious period within the new subpopulation. In addition, although movement of susceptible commuters can lead to the

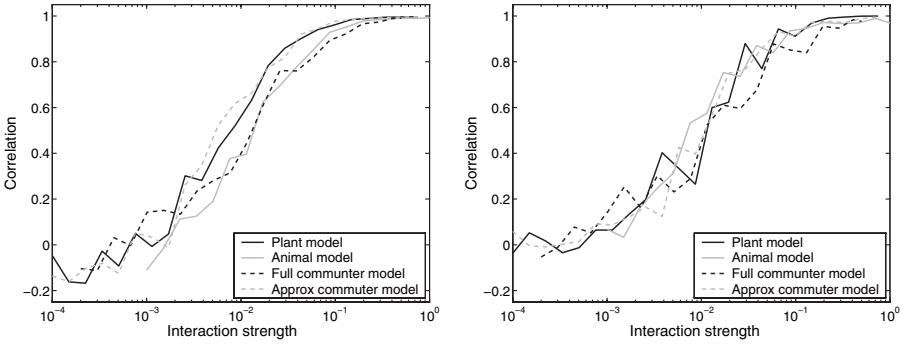


Figure 7.3. The correlation between the levels of an endemic disease in two stochastic populations for a range of interaction strengths and a variety of forms of interaction. It is clear that the correlation increases with the strength of interaction, and eventually asymptotes to one—when the population dynamics are synchronized. For the plant model (equation (7.6), solid black line) and the approximate commuter model (equation (7.10), dashed gray line), the x-axis is the coupling strength ρ. For the animal model (equation (7.8), solid gray line) the x-axis is $m/(\gamma + m)$ whereas, for the full commuter model (equation (7.9) dashed black line, $r_{ij} = r_{ji} = 1$), the x-axis is $2(p)(1 - p)$, where $p = l/(l + r)$, which using equation (7.11) is equal to ρ for the approximate commuter model. ($\gamma = 0.1$, $\beta = 1$, stochastic imports at a rate $\varepsilon = 10^{-2}$, $\mu = 10^{-4}$, left-hand graph $\nu = 1 \Rightarrow N_i = 10^4$, right-hand graph $\nu = 10 \Rightarrow N_i = 10^5$).

transfer of infection between populations—because they can get infected while away and bring the disease home—leading to an additional factor of 2, the same is not true for the permanent migration of animals. Finally, although R_0 increases with coupling in the plant model (equation (7.7)) this effect is not strong enough to affect the correlation between populations.

The correlation between disease prevalence in two subpopulations increases sigmoidally with the strength of interaction between the populations. In general, the change from largely independent dynamics to synchrony occurs for interaction strengths from 10^{-3} to 0.1.

7.2.3. Extinction and Rescue Effects

Fundamental to the dynamics of any spatially segregated population is the pattern of localized extinction and subsequent colonization by rescue events (Foley et al. 1999; Boots and Sasaki 2002; Onstad and Kornkven 1992). As discussed in Chapter 6, smaller populations suffer a greater risk of extinction—intuitively, smaller populations will have fewer infected individuals and therefore be more prone to stochastic individual-level effects. In particular, the extinction risk has been seen to decrease exponentially with population size (Chapter 6; Bartlett 1957, 1960). This has clear implications for spatially segregated populations (and spatial models) where one large population has been subdivided into many smaller ones, with each small subpopulation facing a far greater risk of extinction. Therefore, in a metapopulation where there is no interaction or coupling, the disease in each of the small subpopulations will be rapidly driven extinct, leading to much

swifter global eradication than if coupling is large and the metapopulation was effectively randomly mixed. If we assume that the rate of extinction is proportional to $\exp(-\epsilon N)$, where N is the total population size, then the average time to extinction for the completely mixed (large coupling) metapopulation is:

$$\text{Time to extinction} = \frac{k}{e^{-\epsilon N}} = ke^{\epsilon N},$$

where k is the proportionality constant. In contrast, if the population is divided into n independent, noninteracting subpopulations (zero coupling), then the time to global extinction (in all subpopulations) becomes:

$$\text{Time to extinction} = \frac{k}{ne^{-\epsilon N/n}} + \frac{k}{(n-1)e^{-\epsilon N/n}} + \cdots + \frac{k}{e^{-\epsilon N/n}},$$

$$< k(1+\log(n))e^{\epsilon N/n},$$

which for persistent diseases ($e^{\epsilon N}$ much greater than 1) is far shorter as n becomes large. The above formula comes from calculating the average time to the first extinction when n subpopulations are infected, followed by the average time to the next extinction given that now only $n-1$ subpopulations are infected, proceeding iteratively until all populations are disease free.

Without interaction between the subpopulations, a spatially segregated metapopulation generally suffers a faster rate of stochastic extinction than its randomly mixed counterpart.

When the subpopulations interact, the behavior is far more complex. The transmission of infection to a disease-free subpopulation—termed a rescue event in the metapopulation literature (Hanski and Gilpin 1991, 1997; Hanski and Goggiotti 2004)—can allow the subpopulation to recover from extinction. In this way, long-term persistence can be achieved because infection is constantly reinvading subpopulations that have gone extinct. However, two antagonistic forces are in operation. Rescue events are clearly maximized when the rate that infection enters a subpopulation is large. Intuitively this requires that the interaction or coupling between the subpopulations is large. However, it also requires the populations to be asynchronous, such that when the disease is extinct in one population some of the others have ample infectious individuals; this occurs only when the interaction is fairly weak. As a result, tension exists between high levels of interaction but maintaining asynchrony, and therefore persistence can be maximized at an intermediate level of coupling. A comprehensive understanding of this problem has still not been reached. Keeling (2000b) and Hagenaars et al. (2004) have studied the level of stochastic extinction in two different models of disease metapopulations. The basic structure of the models is very similar, with only minor differences in the mechanism of spatial coupling; however, the two models are focused on diseases with very different properties and time scales. Keeling (2000b) focuses on acute infections comparable to measles, whereas Hagenaars et al. (2004) consider more persistent infections with longer infectious periods and lower R_0. In Figure 7.4, we show examples of the similarities and differences in model behavior reported in these two papers. Both models predict that, at the local scale, higher levels of coupling lead to the disease being present in the subpopulations for longer (less time disease-free, as shown in the top-left graph). However, at the global

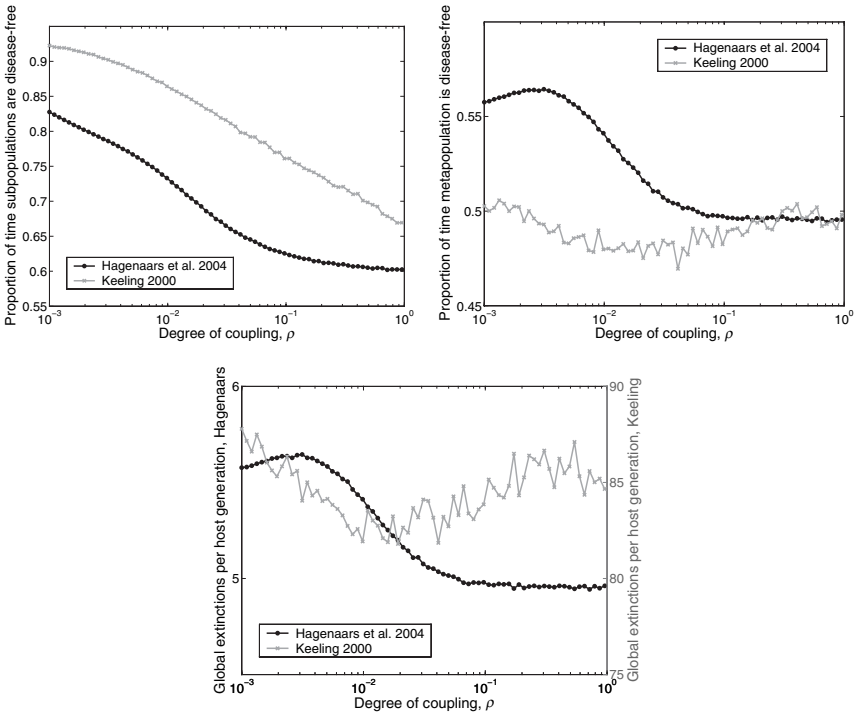


Figure 7.4. Comparison of how the extinction patterns change with the level of coupling from two models: Keeling (2000b) in gray and Hagenaars et al. (2004) in black. The top-left graph shows the proportion of the time that a subpopulation is disease free, whereas the top-right graph shows the proportion of the time that the entire metapopulation is disease free. The bottom graph shows the number of discrete global extinction events (the number of times the entire metapopulation becomes disease free) per host generation; note that the results from the two models are plotted at very different scales. The model of Hagenaars et al. (2004) deals with a small total population size ($N = 10{,}000$), with a relatively protracted infection ($1/\gamma = 125$ days, $R_0 = 5$) with infrequent imports ($\delta = 5 \times 10^{-5}$ day^{-1}; see Section 6.3.3.1), whereas the model of Keeling (2000b) the population size is slightly larger ($N = 30{,}000$), and the disease parameters are measles-like ($1/\gamma = 13$ days, $R_0 = 17$, $\delta = 9.3 \times 10^{-3}$ day^{-1}). In both models the birth and death rates are ($\nu = \mu N$, $\mu = 5 \times 10^{-5}$ day^{-1}), where the other parameters of the Hagenaars model have been scaled accordingly. In both models the population is split into 10 subpopulations.

scale, the two model predictions diverge. The model of Keeling (2000b) shows very little variation in the proportion of time the entire metapopulation is disease free and has a shallow minimum point at $\rho \approx 0.02$. In contrast, the model of Hagenaars et al. (2004) shows more variation and a clear minimum when the coupling is large, $\rho = 1$. Finally, when we focus on the number of global extinction events the Keeling model has again an interior minimum, whereas for the Hagenaars model extinctions are minimized at $\rho = 1$. In addition, extinctions are far more likely in the Keeling model. It is still a major

challenge to incorporate such results into a general framework, understanding how disease characteristics and spatial coupling interact to determine persistence.

When interaction between the subpopulations is included, the level of local (subpopulation scale) and global (metapopulation wide) extinctions is an emergent property of the dynamics and cannot be easily predicted from the disease parameters.

Rather than being an abstract concept, this problem has strong public health implications. The last century has seen far more national and international travel, increasing the coupling between populations (Grais et al. 2003; Cliff and Haggett 2004)—the implications of this for disease persistence and eradication needs to be studied using the form of metapopulation model developed here. Much more work is still needed to determine how population structure, disease dynamics, population size, infectious import rate, and coupling interact to determine disease persistence.

Two other applied problems are strongly related to the issues of coupling, synchrony and extinction. It has long been realized in conservation biology that increasing the mixing between populations, by introducing linking corridors of suitable habitat, is an effective means of conserving a species (Hanski 1999; Earn et al. 2000). However, these corridors may also act as a conduit for the spread of infection, and when facing a highly virulent pathogen, the presence of corridors may exacerbate the extinction of the host (Hess 1996). More recent, detailed work has called into question the strength of this result (Gog et al. 2002; McCallum and Dobson 2002), but the concept is still one that conservationists should consider.

A second applied problem concerns the dynamics of vaccinated metapopulations. Vaccination reduces the prevalence of infection within a population (Chapter 8), which increases its risk of extinction, and also reduces the effective interaction strength because less transmission of infection can occur. This reduction in the effective coupling in turn leads to less synchrony in the disease dynamics and therefore more effective rescue events (Earn et al. 1998; Rohani et al. 1999). This behavior is seen to some degree in the measles data set for England and Wales, where despite vaccinating at around 60% from 1970 to the mid-1980s the rates of stochastic extinction remained unchanged (Keeling 1997). Therefore, the effects of lower infection levels but greater asynchrony of epidemics effectively cancel. Pulsed vaccination (see Chapter 8) has been suggested as a mechanism of overcoming this problem where national vaccination campaigns for a few weeks each year will help to synchronize the dynamics and hence weaken the effect of rescue events (Agur et al. 1993; Nokes and Swinton 1997; Shulgin et al. 1998; Earn et al. 1998). Figure 7.5 shows the correlation (which measures the degree of synchrony) between two coupled populations under standard and pulse vaccination; clearly pulsed vaccination maintains the synchrony of epidemic behavior over of wide range of vaccination coverage.

Pulsed vaccination campaigns act to synchronize epidemics in coupled populations, and may lead to an increase in global extinction rates.

Extending these theoretical vaccination results into a practical public health tool is a challenge for the future. Pulsed vaccination clearly has the ability to synchronize epidemic behavior, which in turn limits the potential of rescue effects between coupled populations. It is hoped that such synchronization may increase the level of global extinctions at the metapopulation level and therefore promote the eradication of infection. Pulsed vaccination can also have significant logistical benefits when individual health care is limited, because

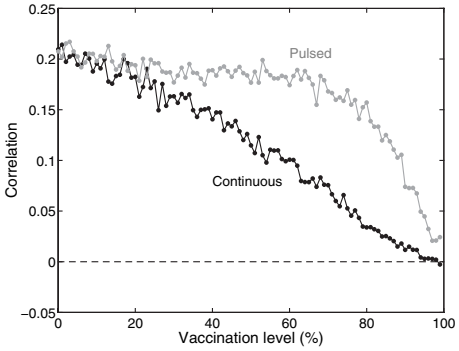


Figure 7.5. The correlation between two coupled populations with SIR-type infection as the level of vaccination is varied. For continuous vaccination (black), individuals are vaccinated at birth, whereas for the pulsed vaccination (gray) a comparable number of individuals are vaccinated every four years. Vaccination is assumed to offer lifelong protection. ($N = 100,000$, $\rho = 0.01$ with approximate commuter coupling, $1/\gamma = 10$ days^{-1}, $R_0 = 10$, stochastic imports $\delta = 5$ per year; see Section 6.3.3.1).

dedicated vaccination teams can be deployed. However, in the delay between the pulses there is the potential for many individuals to become infected. Clearly there is a trade-off: Frequent pulses (or continuous vaccination) limit the buildup of susceptibles, whereas infrequent, and therefore large, pulses have a greater synchronizing action. Determining the appropriate timing of vaccination pulses (or even a mixture of continuous and pulsed vaccination) requires the use of well-parameterized stochastic spatial models that also include age-structure—although the construction of such models can be pieced together from this chapter as well as Chapters 6 and 3, the parameterization and analysis of the possible control permutations is beyond the scope of this book.

7.2.4. Levins-Type Metapopulations

In the metapopulation models considered so far, the population levels within each subpopulation have been modeled explicitly—this may be computationally intensive. An alternative formulation was proposed by Levins (1969) where each subpopulation is simply defined as being either empty (disease-free) or occupied (having infection). There are clear parallels between this new classification of subpopulations, which ignores the precise prevalence and the traditional SIR classification of hosts, which ignores the level of pathogen. The intuitive way to conceptualize Levins-type metapopulations is to assume that localized extinctions and successful recolonization events are extremely rare, so that each subpopulation spends the vast majority of its time either disease free or close to the endemic equilibrium.

If we have a large number of subpopulations, and the coupling is global (so that each subpopulation has an equal probability of reinfecting any other subpopulations), then the probability that a subpopulation is infected, P, is given by:

$$\frac{dP}{dt} = \rho(1-P)P - eP, \tag{7.12}$$

where ρ measures the reinfection (coupling) rate from an infected subpopulation to an uninfected one, and e is the rate of local extinction. Equation (7.12) is structurally identical to the SIS equation (Chapter 2), reflecting the fact that after a localized extinction the subpopulation is once again susceptible to infection.

When the subpopulations are of different sizes and the interactions are unequal, the above formulation can be refined such that P_i now refers to the probability that subpopulation i is infected.

$$\frac{dP_i}{dt} = \sum_j \rho_{ij}(1 - P_i)P_j - e_i P_i, \tag{7.13}$$

where different couplings can occur between different populations and the extinction rate (e_i) can also vary, often reflecting the population size. These rates can easily be used to translate this differential equation model into a stochastic one, with P_i being either zero or one.

Levins metapopulation models ignore the internal dynamics within each subpopulation, and instead classify each subpopulation as either infected or disease free.

Although the Levins formulation is intuitively appealing, the accuracy of the results is highly dependent on the assumption that extinction and successful recolonization events are rare compared to the standard epidemiological dynamics (Keeling 2000b). When this assumption breaks down, two confounding factors associated with the internal subpopulation dynamics become important. First, the conditions that lead to an extinction are unlikely to allow an immediate successful reinfection of the subpopulation. Second, following the extinction of infection the level of susceptibles increases; when there is a high proportion of susceptible individuals any subsequent infected is likely to be large and short-lived, rapidly returning to the disease-free state. Therefore, the timing between extinction and recolonization is vital (see Chapter 6, Figure 6.8). However, these discrepancies between the Levins approximation and the full stochastic metapopulation behavior are often during the early colonization dynamics, making the Levins models an ideal tool to study the spatio-temporal invasion of infection.

Despite differences between the equilibrium-level results of Levins and full metapopulation models, the Levins model still remains a useful and simple tool for studying invasion dynamics.

7.2.5. Application to the Spread of Wildlife Infections

We contrast the Levins metapopulation model with the full metapopulation model by considering two examples of invading wildlife diseases that have been tackled by the two differing approaches. The full metapopulation model of Swinton et al. (1998) was used to investigate the spread of phocine distemper virus around the North Sea coastline. Similar formulations have been used to describe the spread of bovine tuberculosis in badgers (White and Harris 1995), parapoxvirus in squirrels using a grid of stochastic subpopulations (Rushton et al. 2000), and rabies in foxes again using a grid of stochastic subpopulations (Tischendorf et al. 1998). The Levins metapopulation model of Smith et al. (2002) was used to explain the spread of rabies in the raccoon population of Connecticut.

7.2.5.1. Phocine Distemper Virus

In 1988, an epidemic of phocine distemper devastated the harbour seal (*Phoca vitulina*) populations in the North Sea (Dietz et al. 1989). Starting in Anholt, Denmark, in early April, the disease spread in a wave-like manner around the North Sea coastline, triggering epidemics from Norway to Ireland. The UK populations were the last to lose the disease in August of 1989, although the bulk of the cases occurred before the end of 1988.

Swinton et al. (1998) modeled this epidemic as a set of 25 subpopulations, mimicking the 25 locations (seal colonies) where infection was recorded. The dynamics of phocine distemper in this metapopulation were modeled as:

$$\begin{aligned}
\frac{dX_i}{dt} &= -\beta X_i \left[(1-\rho)\frac{Y_i}{N_i} + \rho \frac{\sum_{j=i-1,\,i,\,i+1} Y_j}{\sum_{j=i-1,\,i,\,i+1} N_j} \right] - \mu X_i, \\
\frac{dW_i}{dt} &= \beta X_i \left[(1-\rho)\frac{Y_i}{N_i} + \rho \frac{\sum_{j=i-1,\,i,\,i+1} Y_j}{\sum_{j=i-1,\,i,\,i+1} N_j} \right] - \sigma W_i - \mu W_i, \\
\frac{dY_i}{dt} &= \rho W_i - \gamma Y_i - m Y_i - \mu Y_i, \\
\frac{dZ_i}{dt} &= \gamma Y_i - \mu Z_i,
\end{aligned} \tag{7.14}$$

where $\frac{m}{m+\gamma} \approx 0.2$ gives the probability of mortality from the infection. Swinton et al. simulated this model stochastically and were able to estimate the majority of the parameters needed from good observational data; however, the coupling parameter, ρ, could be found only by matching the model to the observed wavespeed, leading to $\rho \approx 0.1$. The precise form of coupling used in these equations differs from the standard assumptions and therefore requires some explanation. The first term reflects that a fraction $(1-\rho)$ of infectious seals remain at their haul-out site and can therefore infect susceptible seals—the transmission is assumed to be frequency dependent, due to the types of interaction that occur at haul-out sites. The second transmission term is due to encounters with seals at sea, away from the main haul-out sites; again, frequency-dependent transmission is assumed. Therefore, although the formulation of the transmission terms is unusual it reflects the form of interaction between seals. Figure 7.6 shows the wavelike progress of the invading infection; because this is a stochastic model, no two simulations will be identical but the general wave-speed is largely invariant.

Two difficulties exist with this model formulation. First, all subpopulations (seal colonies) are given identical parameters and population sizes. Therefore, heterogeneities that may account for epidemiologically important factors, such as the prolonged epidemic observed in Tayside, are ignored. Second, the model is essentially one-dimensional with nearest-neighbor coupling (see Section 7.3.1); therefore, the true spatial structure has been neglected and the different distances between seal colonies ignored. However, despite these limitations, the model provides a simple means of assessing the spatio-temporal dynamics of phocine distemper epidemics.

7.2.5.2. Rabies in Raccoons

The spread of rabies has been extensively studied using spatial models building upon the pioneering work of Murray et al. (1986), which formulated PDE models (see Section 7.4)

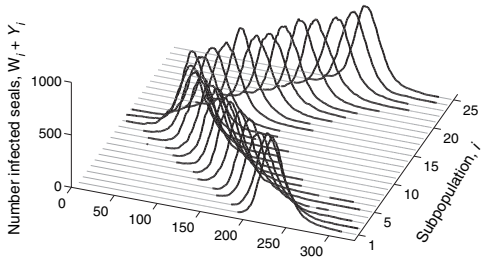


Figure 7.6. Results of a stochastic simulation of the phocine distemper metapopulation model, equation (7.14), as proposed by Swinton et al. 1998. The number of infected seals in each subpopulation is shown, with black lines used when at least one seal is infected. In this model births and deaths have been ignored for simplicity. ($\beta = 0.4$, $\rho = 0.1$, $\sigma = 0.1428$, $\gamma = 0.1143$, $m = 0.0286$, $\mu = 0$, $N_i(0) = 2000$).

to understand the spread of rabies in the fox population of Europe. At a smaller scale, Smith et al. (2002) used a Levins-type metapopulation model to study the spread through Connecticut of rabies in racoons from 1991–1996, concentrating on the underlying spatial heterogeneity of the habitat. This spread is part of a larger wave of infection that began along the Virginia/West Virginia border in the mid-1970s.

Using a similar format as equation (7.13), Smith et al. model the state of the 169 townships in Connecticut, and for township i define the stochastic rate of infection as

$$\varepsilon_i(1 - P_i) + \sum_j \rho_{ij}(1 - P_i)P_j, \tag{7.15}$$

where P_i is one if the racoons in the township are infected and zero otherwise. Unlike the previous Levins metapopulation model (7.13), the localized extinction of infection has been ignored. Here ε_i measures the random long-distance dispersal of rabies due to racoon translocation and ρ_{ij} measures local transmission between adjacent townships (see Figure 7.7). Here $\rho_{ij} = A$ if townships i and j share a land-boundary, $\rho_{ij} = B$ if the townships are separated by a river, otherwise $\rho_{ij} = 0$. This stochastic model (which is an SI model because there is no recovery or local extinction of infection) was compared to the observed first recorded case of rabies in each township, in order to determine suitable coupling and long-range transmission parameters. The simplest model that provides a reasonable fit to the observed data had $\varepsilon_i = 2 \times 10^{-4}$, $A = 0.66$, and $B = 0.09$ (all rates in months), showing that rivers reduce transmission by 87% compared to land boundaries and that local transmission accounts for the vast majority of spatial spread. Other, more complex model formulations included the size of the human population as a proxy for the density of racoons within a township; such models suggest that population density plays a small but positive role in transmission.

Figure 7.8 shows the impact of these three different coupling parameters. Rivers clearly have a significant impact on the spatial spread of infection, especially when long-range translocations are impossible (gray line and middle map). When the impact of rivers is ignored (dashed line and right map), the speed of spatial spread is far more rapid, infecting the entire state in two-thirds of the time. Interesting, although the value of ε is very small, its effect is significant because it breaks the assumption of strict local transmission.

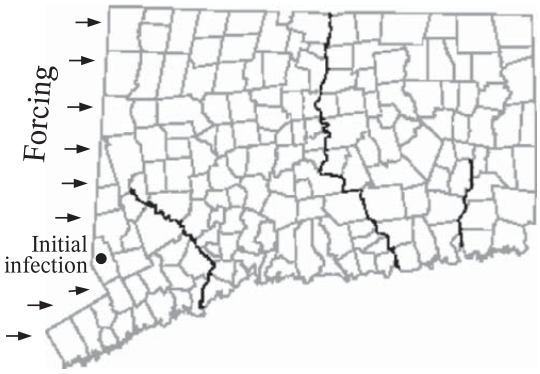


Figure 7.7. The 169 townships within Connecticut showing the initial site of the first recorded cases in Ridgefield Township, the forcing on the western townships due to infection from neighbouring New York, and the position of rivers (black) which act as a natural barrier to the movement of rabid racoons.

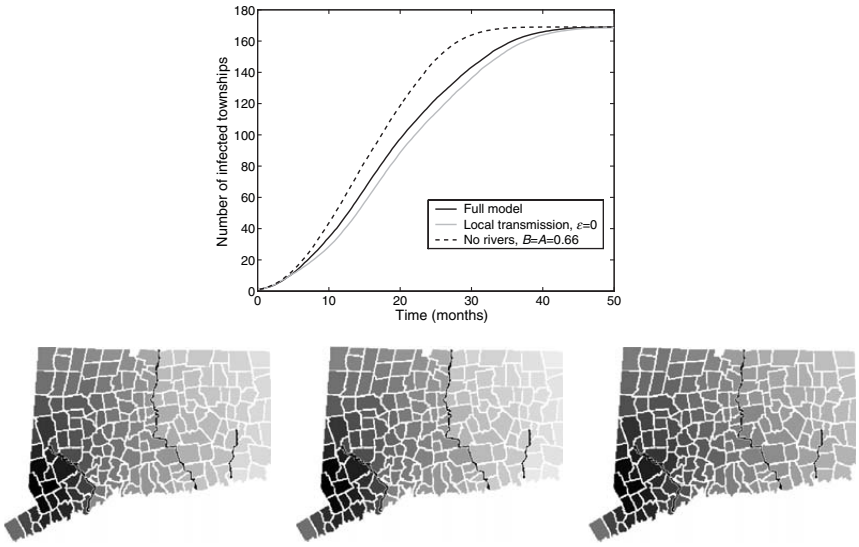


Figure 7.8. The effect of the various coupling terms on the spatial spread of rabies in Connecticut, based on the model and parameterization of Smith et al. (2002). To simplify the dynamics, forcing from New York state is set to zero. The top graph shows the average number of infected townships as a function of time from the initial seeding of infection. Three different assumptions about coupling are shown: the full model ($A = 0.66$, $B = 0.09$, $\varepsilon = 2 \times 10^{-4}$), a model with local transmission only ($A = 0.66$, $B = 0.09$, $\varepsilon = 0$), and a model where rivers had no impact on transmission ($A = 0.66$, $B = 0.66$, $\varepsilon = 2 \times 10^{-4}$). The spatial pattern for these three models is illustrated on the three maps, with darker colors representing earlier average infection times.

By the nature of the Levins-type metapopulation used here, all epidemiological dynamics within a subpopulation are ignored. This means that as soon as a township is infected it can transmit infection as strongly as when endemic equilibrium has been reached. In an ideal situation, the within-subpopulation dynamics should also be included and modeled stochastically. However, the Levins approximation allows for very rapid computation and therefore far richer parameterization; it also circumvents the difficulty of estimating the racoon population levels within each township. This is a good example of where limitations of the available data determine the appropriate model choice, and where simple models can be used to derive the main conclusion that rivers act as significant but permeable barriers to the spread of rabies in racoons.

7.3. LATTICE-BASED MODELS

There are many occasions when the spatial location of the hosts is seen to be important, but there is no natural means of partitioning the entire population into discrete subpopulations. In such situations, lattice- or grid-based models are often used, where individuals within a grid site are grouped together into a subpopulation. Traditionally, these grid-based models take two forms: coupled lattice models that can be considered as a grid of subpopulations with coupling between adjacent grid sites (neighbors), and cellular automata where at most one host can occupy each grid site and again interactions occur only locally with the neighboring sites. In essence, both of these models are special forms of the metapopulation model, with a tightly constrained set of interactions. In general, these models are conceptual tools, used to inform about the effects of spatial separation and nonrandom mixing; they are seldom used as accurate predictive tools, because with few exception (such as orchards) most hosts do not exist on regular lattices. However, the insights provided by this type of model have proved to be invaluable in understanding the spatial spread of infection (White and Harris 1995; Tischendorf et al. 1998; Keeling and Gilligan 2000; Rushton et al. 2000; Kao 2003).

Two main types of lattice exist, which determine how the entire population is divided into subpopulations. Probably the most intuitive is the two-dimension square lattice, so that an individual's position (on the surface of the earth) is translated into x and y coordinates which in turn determine the grid site (White and Harris 1995). Modifications to this standard grid arise by changing the number of dimensions; many theoretical problems can be studied more precisely if the populations are arranged along a one-dimensional line, with interactions between nearest (or nearest and next-nearest) neighbors (Harris 1974; Watts and Strogatz 1998). Alternatively, higher dimensional lattices can also be used to replicate the more complex higher-dimensional social structure of humans (Rhodes et al. 1997). Square lattices impose a definite direction on the space (mathematically we lose the property of isotropy) such that the directions "north," "south," "east," and "west" play a more major role than any others. To overcome this problem, some researchers have used hexagonal grids, where each site has six neighbors; this often generates more natural spatial patterns (van Baalen and Rand 1998; Kao 2003).

7.3.1. Coupled Lattice Models

Although coupled lattice models are simply a special case of the metapopulation formulation, it will be worth considering some of the particular differences in more detail. For

a disease with SIR-type dynamics and "commuter-like" interaction terms (see Section 7.2.1), the governing equations become:

This is online program 7.3

$$\begin{aligned}
\frac{dX_i}{dt} &= \mu - \beta X_i \frac{(1-\sum_j \rho_{ji})Y_i + \sum_j \rho_{ij} Y_j}{(1-\sum_j \rho_{ji})N_i + \sum_j \rho_{ij} N_j} - \nu X_i, \\
\frac{dY_i}{dt} &= \beta X_i \frac{(1-\sum_j \rho_{ji})Y_i + \sum_j \rho_{ij} Y_j}{(1-\sum_j \rho_{ji})N_i + \sum_j \rho_{ij} N_j} - \gamma Y_i - \nu Y_i, \qquad (7.16) \\
\rho_{ij} &= \rho_{ji} = \begin{cases} \rho & \text{if } i \text{ and } j \text{ are neighbors} \\ 0 & \text{otherwise,} \end{cases}
\end{aligned}$$

where i refers to a grid cell within the lattice. (Given that lattices are usually two-dimensional, some researchers prefer to specify a subpopulation by its coordinates; therefore, X_{ij} is the number of susceptibles at location (i, j). However, this notation can be very cumbersome when we wish to specify the level of interaction between two locations (e.g., $\rho_{(i,j)(k,l)}$)).

Equation (7.16) assumes that all populations have identical demographic and epidemiological parameters, and therefore we have moved further away from data-driven models such as those illustrated in Section 7.2.1.3. However, this assumption of homogeneity of parameters allows for a more detailed understanding of the spatio-temporal dynamics. In addition, it is frequently assumed that all populations are of equal size ($N_i = N$), which simplifies the denominator in equation (7.16) to N.

Coupled lattice models are specialized metapopulation models, where subpopulations are arranged on a grid and coupling is generally to the nearest neighbors only.

The most notable feature of this form of lattice model is the clear wave-like spread of invading infections. From a point source of infection, the disease must spread to the neighboring sites before it can spread to the rest of population. Figure 7.9 shows a pictorial example of the wave-like spread for a square lattice, together with a graph of how coupling (ρ) and the basic reproductive ratio ($R_0 = \frac{\beta}{\gamma+\nu}$) both determine the speed of the invading wave.

As with the metapopulation models, a stochastic version of the lattice model has a slower wave speed than the deterministic model, and when the deterministic wave speed is very low the stochastically spreading disease may even fail to colonize the entire lattice. This reduction in speed is because in an integer-based stochastic model transmission of infection into a new subpopulation is a stochastic process and therefore may, by chance, be considerably delayed; by contrast, in a deterministic model very low prevalence will always spread to the new subpopulations. As expected, increasing the level of interaction between neighboring subpopulations, or increasing the within-subpopulation growth rate, allows the infection to spread more rapidly.

The wave speed of an invading epidemic in a coupled-lattice model increases almost linearly with the initial growth rate of the infection, $\beta - \gamma - \nu$; increases nonlinearly with the level of coupling, ρ; and is slightly more rapid in deterministic compared to stochastic models.

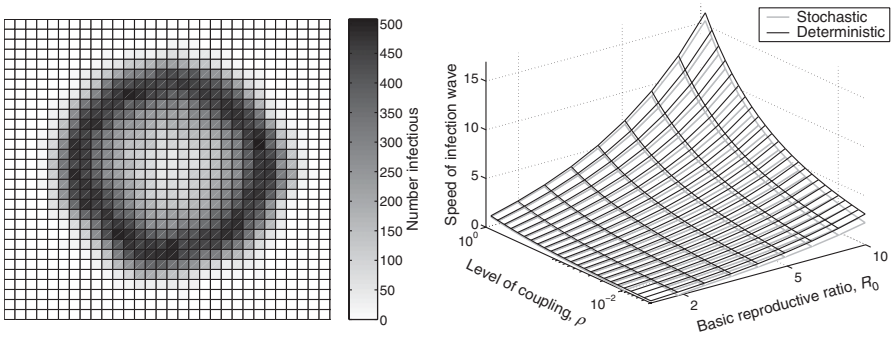


Figure 7.9. Results from a coupled lattice model using the approximate commuter-type coupling (equation (7.16)). The left-hand figure shows a snapshot of the lattice at time $t = 5$, with darker shades illustrating a higher level of infectious individuals ($R_0 = 5$, $\rho = 0.1$). The circular wavelike spread of infection is clear. The right-hand graph shows the speed of the invading wave of infection as the basic reproductive ratio and the level of coupling vary. The model simulates the spread of infection on a 100×100 lattice; the local subpopulations are all identical and contain 1,000 individuals. ($\gamma = 1$; changing this would scale the speed of the invading wave).

In contrast to metapopulations where space is often rather abstract, coupled lattice models provide a definite method of including the spatial location of individuals. This allows us to predict the expected wavelike spread of infection across a homogeneous spatial landscape. As such, coupled lattice models play a vital role in improving our understanding of the spatial spread of infection; they can be readily derived from the standard nonspatial models, and due to the simplistic assumptions about spatial interaction require very little extra information to parameterize. However, such coupled lattice models break down if we attempt to match them too rigorously to the underlying individual-level spatial dynamics. This difficulty arises because of the assumption that mixing within a grid site is completely random (and full strength), whereas mixing between adjacent cells is far weaker. Figure 7.10 shows an example of how the coupled-lattice assumption can fail to capture the expected individual-level behavior. Because individuals A and B are within the same grid cell, the coupled lattice assumption means that the interaction between them is strong, whereas the interaction between B and C (and A and D) is far weaker or zero because they lie in separate cells despite the fact that the separations involved are smaller. This highlights a fundamental flaw with coupled lattice models: The act of artificially aggregating populations into grids can lead to some artificial results and hence the grid size must be chosen with extreme care.

7.3.2. Cellular Automata

Cellular automata also use a lattice-based arrangement of sites. However, in contrast to the lattice-based models discussed above, cellular automata have only a finite, and usually small, number of population states. Most frequently we consider each lattice site to represent a single host (a population size of one) and so each site is generally either empty,

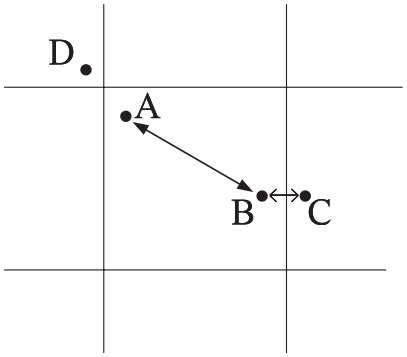


Figure 7.10. A representation of when the coupled-lattice framework may fail to describe the expected interaction between individuals. Although A and D and (B and C) are close, the interaction between them is weak because they occupy different grid locations.

or occupied by a susceptible, infectious, or recovered individual. Due to this finite nature almost all cellular automata disease models are stochastic.

Cellular automata are clearly an abstraction of reality. Other than in agricultural settings (Maddison et al. 1996; Klecakowski et al. 1997; Gibson 1997b), individuals do not often exist in fixed lattice arrangements. In addition, the small number of interaction neighbors that are usually assumed (just the nearest 4 or 8 lattice sites) are unrepresentative of the complex and heterogeneous contacts through which human and animal infections pass. However, cellular automata are superb tools to understand how the individual and spatial nature of populations causes epidemic dynamics to deviate from their deterministic random-mixing ideal.

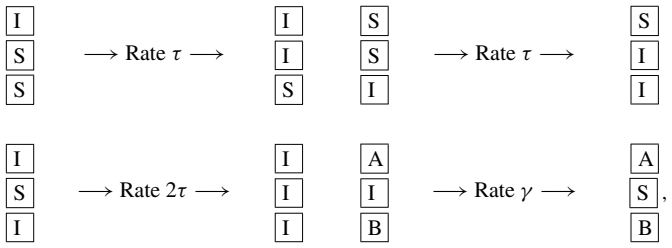
Cellular automata operate on a lattice of sites, with each site generally assumed to hold a single host. Interaction is usually stochastic and with the neighboring (4 or 8) lattice sites.

Two basic forms of cellular automata dominate the early literature in this subject area, and have clear parallels to disease models. These are the contact process, which is equivalent to the SIS model, and the forest-fire model, which is equivalent to an SIR-type infection.

7.3.2.1. The Contact Process

The contact process, which dates back to 1974, is traditionally formulated in one dimension so that individuals are positioned in a row with contact between adjacent individuals (Harris 1974). The stochastic rules that govern the behavior of this model, when translated into a disease metaphor are biologically intuitive. Infectious individuals transmit infection at a rate τ to any neighboring susceptible and infectious individuals recover at a rate γ, becoming susceptible to the disease once more—this is clearly the SIS model (Chapter 2) in a spatial context. By considering a central individual and both neighbors, we can

explicitly define the possible rates of change:



where A and B can be either S or I.

For this contact process we want to calculate R_0, thereby linking the individual level parameters τ and γ to population-level dynamics. If we consider R_0 as the rate that secondary cases are initially produced, multiplied by the infectious period (Chapter 2), then it is clear that $R_0 = 2\frac{\tau}{\gamma}$, where 2 comes from the fact that every individual has 2 contacts. This would suggest that the disease can invade and expand throughout the population if $\tau > \frac{1}{2}\gamma$. However, such arguments do not take into account the spatial structure that develops during the early epidemic process. Consider the initial seed infection, where by chance it infects a neighboring contact; however, this neighbor now has only one susceptible neighbor (unless the seed infection recovers) and therefore has a much reduced potential for spreading the infection. We therefore observe a common phenomenon in many spatial disease models: The local pool of susceptibles is rapidly depleted, which can dramatically reduce the early growth rate.

In many spatial models, the depletion of the locally available susceptible population can reduce the early growth rate of the epidemic and the speed of the invading wave front.

Even for this most simple of individual-based spatial models, the precise threshold value of τ that allows an epidemic to spread can be calculated only by repeated large-scale numerical simulation. The latest estimates suggest that $\tau > 1.64896\gamma$ for the infection to have a nonzero probability of long-term spread and persistence (de Mendonça 1999).

7.3.2.2. The Forest-Fire Model

A second common form of cellular automata is the forest-fire model, which is closely associated with spread of $SIRS$-type infection and is usually simulated on a two-dimensional lattice. In the original notation, lattice sites can be empty, occupied by a healthy tree, or occupied by a burning tree. Burning trees die to leave empty spaces, fire can spread between neighboring trees, trees can colonize empty spaces, and occasional random lightning strikes can cause spontaneous fires. In epidemiological notation, healthy trees are susceptibles, burning trees are infectious, empty sites are recovered (and immune), colonization by trees mimics either the birth of new susceptibles or waning immunity, and lightning represents the import of infection. Again we describe the dynamics in terms of

the rates of change of lattice sites:

$\boxed{S} \longrightarrow \boxed{I}$ Rate = $\tau n + \varepsilon$ where n is the number of infectious neighbors.

This is online program 7.4

$\boxed{I} \longrightarrow \boxed{R}$ Rate = γ.

$\boxed{R} \longrightarrow \boxed{S}$ Rate = ν.

This model was developed by Per Bak and coworkers (Bak et al. 1990), and for statistical physicists displays a range of interesting power-law scaling and self-organized critical behavior. In general, distributions such as the size of patches of susceptibles (trees) or the size of individual epidemics are observed to follow a power-law relationship (frequency $\propto$ size$^{-\alpha}$), with the power-law exponent, α, being largely independent of the precise parameter values. This behavior occurs whenever certain rates of change are much bigger than others; in particular, transmission is much faster than recovery, which is much faster than births, which is much faster than random imports of infection ($\tau \gg \gamma \gg \nu \gg \varepsilon$). Fortunately, this natural ordering holds in most epidemiological examples, so it is hoped that the same power-law scaling and ideas of self-organized criticality will hold also. Much more information on self-organized criticality can be found in the following publications: Tang and Bak (1988); Sole et al. (1999); Allen et al. (2001); Pascual and Guichard (2005).

The forest-fire model typifies many stochastic spatial models. The fact that transmission is faster than recovery, which is faster than births, which is faster than imports of infection, leads to power-law relationships between epidemic size and frequency.

7.3.2.3. Application: Power-Laws in Childhood Epidemic Data

The work of Rhodes and coworkers (Rhodes and Anderson 1997; Rhodes et al. 1997) provides a good example of how cellular automata models can be used to develop deeper insights into the roles of spatial structure and individual-based populations in the dynamics of infectious diseases. They were interested in the dynamics of childhood diseases, and the distribution of outbreak sizes in small isolated populations. The Fareo Islands are isolated islands in the North Atlantic between Scotland and Iceland, and have a population of around 25,000. These island have extremely good historical records of infectious disease outbreaks stretching back for around 100 years (Cliff et al. 1993), and so are an ideal source of data for epidemiological study. In a number of papers, Rhodes and coworkers showed that the outbreaks of childhood diseases in the Faroes follow a power-law relationship, and that a similar scaling can be obtained from a biologically simple cellular automaton model.

In Figure 7.11, we show how the frequency of epidemics greater than or equal to a given size decreases, following a power-law like relationship. For epidemics of fewer than 500 cases, the probability that the epidemic is greater than or equal to size s is given by:

$$\mathbb{P}(\text{epidemic} \geq s) = s^{-\alpha},$$

with $\alpha \approx 0.265$ for measles, $\alpha \approx 0.255$ for whooping cough, and $\alpha \approx 0.447$ for mumps. This is reminiscent of the power-law relationships that are seen in the traditional

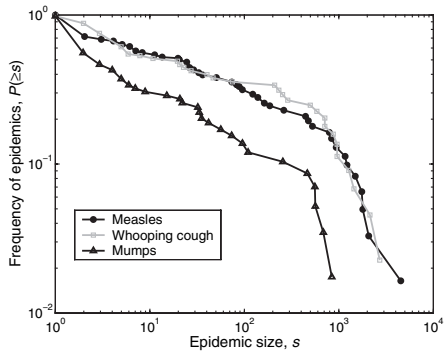


Figure 7.11. The power-law scalings observed in the epidemic outbreak data from the Faroe Islands, 1870–1970. The frequency of epidemics greater than or equal to a given size is shown for three major childhood infections: measles, whooping cough (pertussis), and mumps. It is postulated that a power-law scaling holds for epidemics up to 500–1,000 cases. The graph is plotted on a log-log scale, so that straight lines equate to power-laws. (From Rhodes et al. 1997.)

forest-fire model; therefore, Rhodes and Anderson modified the standard model to overcome some of the limitations and incorporate more realistic human behavior. In particular, the susceptible and infectious hosts are no longer fixed but are able to wander across the lattice by moving into neighbored unoccupied cells. This movement of individuals means that the behavior of adjacent sites is related (because the movement from one cell must be balanced by the movement *to* another), breaking the independence assumptions of formal cellular automata. However, the same basic techniques and results still translate between true cellular automata and this more complex model.

Figure 7.12 shows examples of the Rhodes and Anderson model in two and three dimensions. In two dimensions, the lattice is a 158×158 square, and each site has four nearest neighbors involved in the transmission of infection and movement; in three-dimensions, the lattice is a $29 \times 29 \times 29$ cube, and each site has six neighbors. Although the power-law distribution of (relatively) small epidemic sizes is largely independent of the precise parameter values, the behavior does vary with the dimension of the system. In three dimensions, the distribution of epidemic sizes is a much closer fit to a power-law of $\alpha \approx 0.2$ than when the model is two-dimensional, and therefore seems a plausible representation of measles or whooping cough. When the lattice is made five-dimensional, the power-law changes to $\alpha \approx 0.4$ (Rhodes et al. 1997), and therefore is comparable with the scaling exponent observed for mumps.

It is important to ask what these theoretical results mean at a practical level. Rather that telling us that models in three and five dimensions are required to capture the behavior of measles, whooping cough, and mumps, these results indicate that simple nearest-neighbor transmission in two dimensions is insufficient to describe the true nature of human social contacts. Three and five dimensions are simply methods of introducing the added elements of complexity, clustering, and interaction that exist in human social networks. This reinforces our original contention that most lattice models are abstract tools that can be used to improve our understanding of transmission in a spatial environment, rather than detailed predictive models.

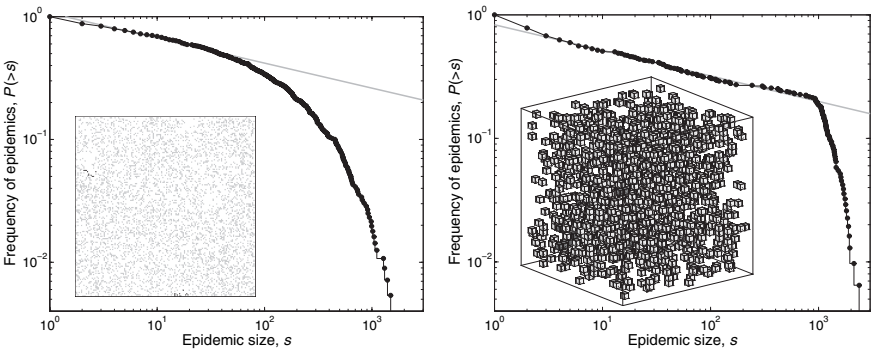


Figure 7.12. Distribution of epidemic sizes and examples of the lattice configuration for 2D and 3D versions of the Rhodes and Anderson model. In the lattice example, gray sites are susceptible, black are infectious, and empty sites are recovered. For both models, the best power-law fit to the distribution of epidemic sizes ($\mathbb{P}(\text{epidemic} \geq s) = s^{-\alpha}$) is calculated to be $\alpha \approx 0.2$. ($\tau = 5$, $\gamma = 0.1$, $\nu = 4 \times 10^{-5}$, imports $\varepsilon = 3.3 \times 10^{-6}$, movement rate $m = 0.1$. $N \approx 25000$.)

7.4. CONTINUOUS-SPACE CONTINUOUS-POPULATION MODELS

One major disadvantage of the lattice-based models is the discretization of space that is introduced by the lattice structure. In consequence, the resolution at which we know an individual's position is limited by the scale of each grid cell. An alternative formulation is to treat both space and the population as continuous, therefore specifying a density of individuals at all locations. We can think of this as the limit of a lattice model as the grid size becomes infinitely fine scale.

The natural way to describe the dynamics of continuous populations in continuous space is using partial differential equations (PDE), or integro-differential equations (IDE). The mathematics behind these formalisms is complex and often highly technical; the details of this approach are excellently described by Murray (2003), so here we review the salient epidemiological implications. Although some notable examples of continuous space models are being used for applied modeling purposes (Noble 1974; Murray et al. 1986; Caraco et al. 2002), such models are generally used to provide theoretical predictions and a generic understanding of the spatial spread of infection (Lopez et al. 1999; Beardmore and Beardmore 2003; Reluga 2004). The main theoretical advantage is the deterministic and tractable nature of the continuous-space models, whereas the assumption of continuous-population levels tends to be a disadvantage in many applied situations. We now focus on the two forms of models, PDEs and IDEs, to illustrate how they are derived from simple nonspatial models (Chapter 2).

7.4.1. Reaction-Diffusion Equations

The standard PDE models are derived from the assumption that infectious individuals transmit the disease only to susceptibles at their current location, and that all individuals are free to move at random (or diffuse) through the landscape. We first show an example of a PDE model, before explaining the individual terms and how such a model is formulated.

A typical PDE model for a disease with SIR-type dynamics would be:

$$\begin{aligned}\frac{\partial X}{\partial t} &= \nu - \beta XY/N - \mu X + D_X \nabla^2 X,\\ \frac{\partial Y}{\partial t} &= \beta XY/N - \gamma Y - \mu Y + D_Y \nabla^2 Y,\\ \frac{\partial Z}{\partial t} &= \gamma Y - \mu Z + D_Z \nabla^2 Z,\end{aligned} \tag{7.17}$$

where X, Y, and Z are functions of both space and time, and represent the local density of susceptible, infectious, and recovered individuals, and as always $N = X + Y + Z$. Hence, if we're dealing with a two-dimensional landscape, $X(x, y, t)$ is the density of susceptibles at location (x, y) at time t. We now have to specify the rates of change with partial derivatives (e.g. $\frac{\partial}{\partial t}$), because our variables are now multi-dimensional, being functions of both space and time—this, however, is a technicality and does not change our understanding of what these derivatives mean.

The term ∇^2 is introduced to model the local diffusion of individuals through space. ∇ is shorthand for the rate of change of the quantity across space, so ∇^2 is the change in the rate of change. In two dimensions, the diffusion term for susceptibles becomes:

$$\nabla^2 X = \frac{\partial^2 X}{\partial x^2} + \frac{\partial^2 X}{\partial y^2}.$$

The inclusion of these spatial derivatives mimics the diffusion of individuals across the environment. For greater generality, susceptible, infectious, and recovered individuals are assumed to diffuse at different rates (D_X, D_Y, and D_Z), reflecting the fact that sick individuals may be less likely to move.

We can understand the role of diffusion in such PDE models, by considering the diffusion of a group of susceptibles initially piled at a single point (0,0). Ignoring demography, our equation will be:

$$\frac{\partial X}{\partial t} = D_X \nabla^2 X.$$

This has the solution:

$$X(x, y, t) \propto \frac{1}{2\pi D_X t} \exp\left(-\frac{(x^2 + y^2)}{2 D_X t}\right),$$

which is an ever-expanding bell-shaped (Gaussian) distribution, with the total density of individuals remaining constant. The diffusion parameter D_X governs the speed at which the variance of the Gaussian grows. We can therefore see that diffusion away from a point source leads to Gaussian-like distributions of individuals.

Reaction-diffusion models, which use a PDE formulism, assume local transmission of infection and rely on spatial diffusion of hosts to spread the infection.

Very few PDE models have an exact analytical solution and therefore numerical methods are required for their simulation. Although PDEs are formulated as continuous space models, determining their solution by computer necessitates discretizing space, usually into a regular grid (see Box 7.1). Therefore, for the vast majority of situations, PDE models are approximated by coupled lattice models with a very fine resolution lattice.

Box 7.1 Solving Diffusion PDEs

Although PDE models are formulated in continuous space, numerical solution of the equations often requires us to discretize space in some manner. The most common method of achieving this is to subdivide space into a regular grid, therefore we are again dealing with a lattice-based model. Here we explain how to translate the differential terms into a lattice formulation.

We impose a lattice structure onto our space, where adjacent lattice points are separated by a distance d; we therefore want the lattice solution $X_{i,j}(t)$ to approximate the PDE solution $X(i \times d,\, j \times d,\, t)$. We are familiar with ways to treat temporal derivatives (e.g., $\frac{d}{dt}$), integrating forward in time using methods such as forward Euler or Runge-Kutta. However, the PDE model also contains spatial derivatives and these need to be expressed in terms of the lattice structure. Considering the x spatial second derivative for the number of susceptibles:

$$\frac{\partial^2 X_{i,j}}{\partial x^2} \approx \frac{\dfrac{\partial X_{i+\frac{1}{2},j}}{\partial x} - \dfrac{\partial X_{i-\frac{1}{2},j}}{\partial x}}{d}.$$

In words, this means the second derivative can be approximated as the change in the first derivative of X between $(i+\frac{1}{2},\, j)$ and $(i-\frac{1}{2},\, j)$ divided by the distance, d, between $(i+\frac{1}{2},\, j)$ and $(i-\frac{1}{2},\, j)$. We now perform a similar approximation for the first derivatives in this term:

$$\begin{aligned}\frac{\partial^2 X_{i,j}}{\partial x^2} &\approx \frac{\left(\frac{X_{i+1,j}-X_{i,j}}{d}\right) - \left(\frac{X_{i,j}-X_{i-1,j}}{d}\right)}{d},\\ &\approx \frac{X_{i+1,j} - 2X_{i,j} + X_{i-1,j}}{d^2}.\end{aligned}$$

Therefore, if we consider the full diffusion term:

$$\begin{aligned}D_X \nabla^2 X_{i,j} &= D_X \frac{\partial^2 X_{i,j}}{\partial x^2} + D_X \frac{\partial^2 X_{i,j}}{\partial y^2},\\ &\approx \frac{D_X}{d^2}\left(X_{i+1,j} + X_{i-1,j} + X_{i,j+1} + X_{i,j-1} - 4X_{i,j}\right).\end{aligned}$$

This spatial approximation can now be substituted in the PDE equations to give an ODE equation for each lattice point, which we can solve in the usual manner. Diffusion therefore acts like the movement of individuals between the four nearest-neighbor lattice sites. The rate of this movement (or coupling) is $\frac{D_X}{d^2}$, such that it is proportional to the diffusion coefficient but increases quadratically with the number of lattice sites that represent one unit length. It is therefore clear that PDEs can be approximated by very fine scale lattices, with very high levels of movement between neighboring sites.

Figure 7.13 gives an example of the types of spatio-temporal dynamics that can be observed using a PDE model for the spread of an SIR-type infection. Starting at a point source, infection spreads as an expanding epidemic wave, leaving secondary oscillations around the endemic equilibrium in its wake. The left-hand graph of Figure 7.13 shows a snapshot of this circular wave front. However, the right-hand graph provides a more intuitive understanding of the spatial pattern, plotting disease prevalence as a function of the distance from the initial source (black solid line). This is compared to the solution of the standard (nonspatial) SIR model (gray dashed line); clearly there is good agreement between these two, although due to the movement of susceptibles into infected regions, the PDE shows a slightly slower decay of the epidemic because diffusion of susceptibles is playing a comparable role to births, allowing infection to persist locally.

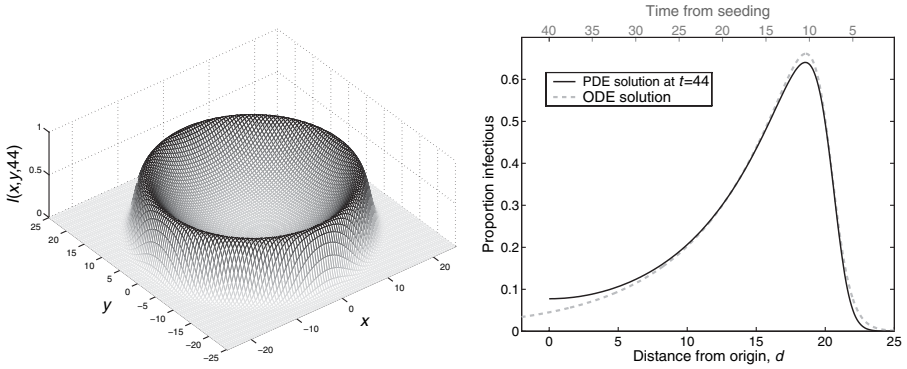


Figure 7.13. Results from the SIR-type PDE model equation (7.18). The left-hand figure is a snapshot, at time $t = 44$, of the density of infecteds in the PDE model. The right-hand figure compares the distribution of infection (at time $t = 44$) as a function of distance from the initial source, with the results from a standard (nonspatial) SIR model with the same basic parameters. For the nonspatial model the x-axis represents the time from the start of the simulation, whereas for the PDE model the x-axis represents the distance from the initial point of infection. The values on the x-axis have been scaled by the wave speed so that the two curves coincide. (The PDE was simulated using a 501×501 lattice, $N = 1$, $\nu = \mu = 10^{-3}$, $\gamma = 0.1$, $\beta = 1$, $D_X = D_Y = D_Z = 0.1$.)

The comparison between the diffusion-based PDE and the nonspatial SIR model hints at a deeper relationship. Involved mathematical calculation shows that, once transient dynamics have died away, the PDE leads to a traveling wave with constant velocity, c (Box 7.2). All such traveling waves (initiated at the origin) can be written as:

$$Y(x, y, t) = \widehat{Y}(r - ct), \qquad \text{where } r = \sqrt{x^2 + y^2}.$$

Therefore, if we stand in one place and record the wave moving past us, we observe the same profile as looking at the spatial wave form at a given time.

For PDE models (in two dimensions), infection spreads as a growing circular wave of near constant velocity.

7.4.2. Integro-Differential Equations

Integro-differential equations (IDEs) share the continuous-space and continuous-population assumptions of the PDE models, but provide far greater flexibility in the way in which infection is transmitted. The PDE model was concerned with the very localized spread of infection and the movement (diffusion) of individuals; in contrast, the IDE models focus on longer-range transmission from static individuals—although this latter constraint is not always true. Although a wide variety of model forms exist, the following

Box 7.2 Speed of Infection Wave

Calculating the invading epidemic wave speed analytically for a given set of diffusion-based PDEs is a complex procedure, which is covered comprehensively and in great detail by Murray (2003). Here we give a brief outline of the concept for the simple epidemic (with no births or deaths) in a two-dimensional population. We start with the basic equations together with host diffusion:

$$\begin{aligned}\frac{\partial X}{\partial t} &= -\beta XY/N + D\nabla^2 X,\\ \frac{\partial Y}{\partial t} &= \beta XY/N - \gamma Y + D\nabla^2 Y.\end{aligned}$$

We now assume that the number of susceptibles and infecteds are circularly distributed ($X(x, y, t) = \hat{X}(r, t)$, $Y(x, y, t) = \hat{Y}(r, t)$ where $r^2 = x^2 + y^2$); this changes the equations to:

$$\begin{aligned}\frac{\partial \hat{X}}{\partial t} &= -\beta \hat{X}\hat{Y}/N + D\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \hat{X}}{\partial r}\right),\\ \frac{\partial \hat{Y}}{\partial t} &= \beta \hat{X}\hat{Y}/N - \gamma \hat{Y} + D\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \hat{Y}}{\partial r}\right).\end{aligned}$$

Next, we assume that the dynamics can be written as a traveling wave solution with velocity c ($\hat{X}(r, t) = \tilde{X}(z)$, $\hat{Y}(r, t) = \tilde{Y}(z)$ where $z = r - ct$):

$$\begin{aligned}-c\frac{d\tilde{X}}{dz} &= -\beta \tilde{X}\tilde{Y}/N + \frac{D}{r}\frac{d\tilde{X}}{dz} + D\frac{d^2\tilde{X}}{dz^2},\\ -c\frac{d\tilde{Y}}{dz} &= \beta \tilde{X}\tilde{Y}/N - \gamma \tilde{Y} + \frac{D}{r}\frac{d\tilde{Y}}{dz} + D\frac{d^2\tilde{Y}}{dz^2}.\end{aligned}$$

Looking at the $\tilde{Y}$ equation, and assuming that r is large so that we are looking for the long-term large-radius dynamics:

$$D\frac{d^2\tilde{Y}}{dz^2} + c\frac{d\tilde{Y}}{dz} + \frac{\beta \tilde{X}\tilde{Y}}{N} - \gamma \tilde{Y} = 0$$

At invasion, when $X = N$, this differential equation only has wavelike solutions when $Ds^2 + cs + (\beta - \gamma) = 0$ has real solutions for s. This provides a lower bound for c; this lower bound is the observed wave speed:

$$c = 2\sqrt{D(\beta - \gamma)} = 2\sqrt{D(R_0 - 1)\gamma}.$$

Therefore, the wave speed is proportional to the square root of the diffusion coefficients, the square root of $(R_0 - 1)$, and the square root of the rate of recovery. This is an *asymptotic* wave speed, which only holds at large radii and once an invading wave front has fully developed.

relatively simple example illustrates the salient points of integro-differential equations:

$$\begin{aligned}\frac{dX(x,t)}{dt} &= \nu(x) - \lambda(x,t)X(x,t) - \mu(x)X(x,t),\\ \frac{dY(x,t)}{dt} &= \lambda(x,t)X(x,t) - \gamma Y(x,t) - \mu(x)Y(x,t),\\ \text{where } \lambda(x,t) &= \beta\int Y(y,t)K(x-y)dy,\end{aligned} \tag{7.18}$$

where the dependence on location, x (which could be in one, two, or more dimensions), and time, t, have been explicitly stated, and the demographic parameters are allowed to vary between locations. The equation for the number of susceptible and infectious individuals is the same as in Chapter 2; it is only through the force of infection, λ, that spatial interactions enter the dynamics.

The force of infection, $\lambda(x,t)$, models the transmission of infection from all points in space (labeled y in the integral) to the point x that we are considering. The transmission rate is assumed to vary with the distance between the susceptible and infectious individual $(x-y)$, and is described by a transmission kernel, K. In simple terms, K defines how infectivity decreases with distance. Clearly, this gives far greater flexibility than achieved by the PDE model; rather than transmission being a local event, it can now occur over a variety of scales. Equation (7.18) models transmission as a density-dependent process; making transmission frequency dependent is more complex because it depends on how we expect the transmission to operate. Two possible alternatives are:

$$\lambda_1(x,t) = \beta \int \frac{Y(y,t)}{N(y,t)} K(x-y)dy, \qquad \lambda_2(x,t) = \beta \frac{\int Y(y,t)K(x-y)dy}{\int N(y,t)K(x-y)dy}.$$

In the first formulation, it is the local proportion infectious at each point that is important, which most closely mimics the situation where interactions at each point y take place sequentially, so that transmission is based on the point frequency. In contrast, the second formulation corresponds to simultaneous interaction with individuals from a range of points, such that it is the averaged proportion infectious that is important.

With integro-differential equations, the spatial spread of infection is via a transmission kernel that defines how transmission risk decays with distance.

Various properties can now be defined for this type of model. Here we will assume that demographic and epidemiological parameters are invariant across space and that space is infinite and two-dimensional, but this does not necessarily have to be the case. To simplify the calculations we rescale the parameters such that the population density at each point is one, $N=1$. First, we consider R_0, which again will predict the likely success of an epidemic.

$$R_0 = \beta \int_{\mathbb{R}^2} K(y)dy = 2\pi\beta \int_0^\infty rK(r)dr.$$

Therefore, for the basic reproductive ratio to be finite we require that the kernel, $K(r)$, eventually decreases faster than r^{-2}. If the kernel decays more slowly (has a "fat tail"), then the integral is infinite; it is difficult to envisage situations where this is a reasonable assumption. In a similar manner, the average dispersal distance, D, is given by:

$$D = \frac{\beta \int_{\mathbb{R}^2} \|y\| K(y)dy}{\beta \int_{\mathbb{R}^2} K(y)dy} = \frac{2\pi\beta}{R_0} \int_0^\infty r^2 K(r)dr,$$

and for this to be finite requires that the tail of $K(r)$ decays faster than r^{-3}. Finally, when the variance of the dispersal distance $var(D)(= \frac{2\pi\beta}{R_0} \int_0^\infty r^3 K(r)dr - D^2)$ is also finite ($K(r)$ decreases faster than r^{-4}), we observe a wave of infection that moves with a constant speed. However, if the variance is infinite but the average is finite ($K(r)$ decays slower than r^{-4} but faster than r^{-3}), the wave front accelerates indefinitely (Diekmann 1978; van den Bosch et al. 1990; Mollison 1991; Shaw 1995). Therefore, Gaussian and

exponential kernel distributions always lead to a traveling infectious wave front moving at a constant speed, whereas power-law kernels can give rise to a wide variety of behavior.

The shape of the tail of the transmission kernel (which defines the long-range transmission) determines the eventual spatial pattern of invasion, from wavelike spread, to scattered local foci, to highly probable extremely long-range jumps.

These results have worrying implications. It is the long-range transmission of infection, defined by the tail of the kernel distribution, that determines the ultimate dynamics of the spatial spread of infection. However, (assuming R_0 is finite) very few secondary cases actually occur at long range and therefore the precise shape of the kernel is difficult to estimate from observational data. Instead, the long-range behavior of the kernel is often estimated only from the population-level properties of the invading wave.

As with the diffusion-based PDE models, very few of these integro-differential equation models have exact analytical solutions, so numerical simulations are required. Again this computational exercise will require the discretization of space into a fine-resolution grid of subpopulations, breaking the continuous space assumption.

7.5. INDIVIDUAL-BASED MODELS

Individual-based models can encompass a wide range of model forms, and can be designed to include a variety of complex and detailed host behavior that could not be readily expressed within the other model types (Nielen et al. 1999; Mangen 2002; Bates et al. 2003; Noordegraaf et al. 2000; Stacey et al. 2004; see also Section 6.3.5). As the name suggests, these models consider the dynamics of individuals that occupy a spatial landscape. In general, these individual-based models have properties in common with both continuous-space models and stochastic metapopulation models; a transmission kernel is generally used to capture the spatial spread of infection but this is tempered by the stochastic, individual-based nature of the population processes leading to a slower rate of spatial spread (Lewis 2000). Here, as an example of this methodology, we will formulate a general stochastic individual-based model where each host is capable of localized movement and transmission is distance dependent; this gives rise to five distinct probabilistic events: transmission, recovery, birth, death, and movement.

Transmission. Transmission is captured using a technique similar to the integro-differential equations models. The rate of transmission (or force of infection) to a susceptible individual, i, is given by:

$$\lambda_i = \beta \sum_{j \in \text{infectious}} K_T(d_{ij}),$$

where d_{ij} is the distance between the susceptible individual i and an infectious individual j; K_T is the transmission kernel that measures how transmission decreases with distance.

Recovery. As is common in almost all the models of this book, the recovery of an infectious host is independent of its environment. Therefore, the recovery rate of infectious individual j is constant: $G_j = \gamma$.

Birth. For many species, the birth rate is a function of two local density components. First, there must be sufficiently many others in the local environment so that the individual can

find a mate. Second, fecundity is generally a function of an individual's fitness and hence decreases with the amount of competition it suffers—a further two kernels may therefore be required.

Death. The natural demographic processes will depend on the particular host species under consideration—clearly wildlife and human populations behave very differently. However, we can safely assume that in many situations the death rate (D_i) of an individual will be a function of the local density, which in turn is a measure of the local competitive effects—this requires a second kernel to calculate the strength of competition.

Movement. Finally, individuals are allowed to move through the environment. The simplest assumption is that individuals randomly and spontaneously move from their current location to a new location given by a local movement kernel. For many species movement may well be more complex, with both the rate of movement and the choice of new location dependent upon the strength of local competition.

Individual-based models account for the spatial interaction between individual hosts distributed on a spatial landscape. They can include a wide variety of complex (more biologically realistic) behavior that often features a spatial component; this can often lead to a huge number of parameters that can be difficult to determine from available data.

Although generic individual-based models such as the one described above are easy to formulate and intuitively appealing, their parameterization may be far more difficult. In the above example, up to five different spatial kernels require estimation. Understanding how errors and biases in these estimates affect the epidemiological dynamics is a computationally intensive problem, but one that cannot be neglected if models are to be robust. In the two model examples given below, only one kernel is required, because there is no demography or movement of individuals, which greatly simplifies both the parameter estimation and the simulation of the spatial epidemics.

One potential difficulty with the simulation of individual-based models is the number of interaction terms that must be considered when the population size becomes large. In principle, these models must consider the probability of each infectious individual infecting each susceptible individual—with the number of permutations being vast. Consider trying to run a stochastic spatial individual-based model with a population of N individuals; the number of infectious-susceptible combinations grows proportional to N^2, whereas the time step between events decreases like N^{-1} (see Chapter 6); therefore, the computational time needed to simulate an epidemic grows proportional to N^3. This means that it becomes increasingly prohibitive to model very large populations. Some approximations and computational shortcuts can be used to dramatically reduce the simulation time and are worth considering in any practical application; these are outlined in Box 7.3, although the techniques given in Section 6.3.5 may also be beneficial.

This is online program 7.5

7.5.1. Application: Spatial Spread of Citrus Tristeza Virus

The model of spatial spread used in this example bridges the gap between cellular automata and individual-based models. We focus on the spread of Citrus Tristeza virus (CTV) between trees in an orchard. Individual trees in an orchard are automatically in a lattice-like distribution as used in cellular automata models. However, transmission is modeled by a transmission kernel and is not confined to simple nearest-neighbor interactions.

Box 7.3 Short-Cuts for Individual-Based Models

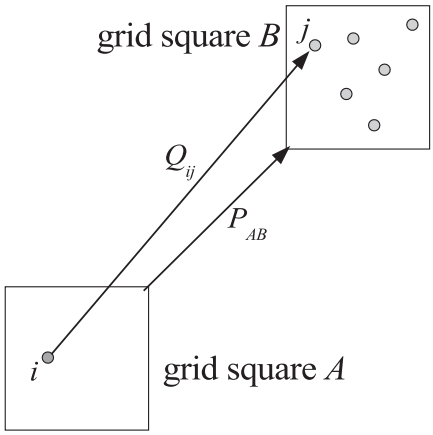
Two main shortcuts are used to improve the speed of individual-based models:

Discrete time. One difficulty with individual-based models is that as the population size becomes larger, the time interval been events decreases, which in turn increases the number of iterations that must be performed to advance the model by a given time period. By discretizing time and allowing multiple events to occur in each time-step, a great improvement in speed is achieved—this is effectively the τ-leap model discussed in Box 6.5. This approximation can be justified for two reasons. First, in a short time-step the chance of two events occurring in a local neighborhood is small, therefore it is likely that all events that occur in one time-step are independent. Second, most biological systems have a natural time frame (often one day), and at finer (or continuous) timescales the standard assumption of constant parameters may not be true. For example, human populations have a clear daily cycle, and unless the continuous time model differentiates between night and day, a discrete-time model with a daily time step may be equally justifiable.

In such discrete-time models, it is necessary to convert the event rates into probabilities:

$$\mathbb{P}(\text{event}) = 1 - \exp(-\text{Rate} \times \text{time step}).$$

Discrete space. By overlaying space with a grid of sites (squares) and knowing which individuals belong in each square, substantial computational saving can be made in terms of the interaction between susceptible and infectious individuals. If the infection process is very localized, then the disease is unlikely to spread to very distant squares, so considering all susceptible individuals within a distant square is often a fruitless exercise. Instead, it is more efficient to first consider whether infection spreads from the infected individual to any host in the susceptible square.



Pictorial example of the mechanism behind using gridded data. We first decide whether infected individual i is likely to have transmitted infection into square B, and only if this probably is realized do we consider the individual hosts within square B.

As an example, consider an infectious individual i in square A and X_B susceptible individuals in square B. An upper bound on the chance that at least one susceptible individual is infected is

given by:

$$\mathbb{P}(A \text{ to } B) = 1 - (1 - P_{AB})^{X_B},$$

where P_{AB} is the greatest possible probability of infection between individuals in squares A and B, often calculated from the shortest distance between squares A and B.

Only if this probability of infection is realized is it worth considering the X_B individuals within square B. Of course, when considering these individuals it must be taken into account that the site-level probability has already been realized.

Once it has been determined that infection into square B is realized ($RAND < \mathbb{P}(A \text{ to } B)$), the pseudo-code below shows how individual-level infection should be treated:

```
s = 1
loop j = 1 : X_B
P = 1 − s(1 − P_AB)^(X_B+1−j)
R = RAND
        if R < P_AB / P
                s = 0
                Q_ij = ℙ(jth susceptible individual in square B is infected
                        by infectious individual i in square A)
                if R < Q_ij / P
                        Infect the jth individual in square B.
                end if
        end if
end loop
```

Here the first "if" statement determines whether the individual j in square B would be infected, assuming that the infectious probability is the maximum value P_{AB} and ensuring that at least one individual within the square would be infected if P_{AB} were the infectious probability for all individuals. This statement is therefore the individual-based equivalent of the square-level condition $RAND < \mathbb{P}(A \text{ to } B)$. The second if statement accounts for the fact that P_{AB} is an overestimate of the true transmission probability. This grid-based method confers great time savings because large numbers of susceptibles in distant squares can be dismissed simultaneously. The most efficient square size is generally found by experimentation. If the grid size is too small, then a large number of squares will need to be considered; however, if the grid size is too big, there is a high chance that at least one individual within it is infected, forcing us to consider all individuals within the square. In principle, a hierarchy of grids could be used to overcome this problem, but it is not clear that the computational efficiency would justify the greater complexity.

Citrus Tristeza virus is an infectious disease of citrus trees that is spread by the brown citrus aphid. It has a worldwide distribution, from Asia to Africa to America, and leads to a reduced fruit crop and eventual loss of the tree, depending upon the strain of virus. In fruit-growing areas, CTV can therefore have severe economic consequences and so in Florida ring-culling has been implemented, removing all citrus trees within 1,900 feet of identified infected trees.

In a series of papers, Gibson and coworkers (Gibson and Austin 1996; Gibson 1997a, b) have examined the spread of CTV using sequential data from detailed studies of its spread in an orchard (Marcus et al. 1984); Figure 7.14 shows an example of this data illustrating the spread of infection.

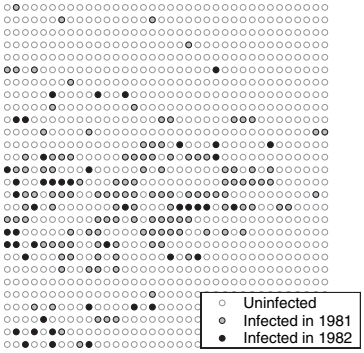


Figure 7.14. Position of infected and susceptible trees in an orchard. Empty circles are susceptible trees, gray-filled circles are trees infected with CTV before the end of 1981, and black-filled circles are trees infected in 1982. Data from Marcus et al. (1984).

In general, the work of Gibson has focused on obtaining accurate parameter values for the spread of CTV infection, which is a highly complex procedure because there are multiple potential sources for each new case. Although detailed parameterization is clearly a prerequisite before model simulation can begin, here we focus primarily on the results from models. CTV is vector transmitted and trees eventually die from infection, however over relatively short spatial and temporal scales we can simply consider the disease as an SI infection (ignoring deaths) and assume that transmission occurs from tree to tree. The proposed model assumes that the rate at which tree i is infected by tree j (assuming i is susceptible and j is infectious) is proportional to a distance kernel $K(d_{ij})$, where d_{ij} is the distance between the two trees. Therefore, the rate that tree i becomes infected (or the force of infection to tree i) is given by the sum over all the rates of infection from all infectious sources:

$$Rate = \lambda_i \propto \sum_{j \in \text{infectious}} K(d_{ij}). \tag{7.19}$$

Gibson argues that a power-law decay ($K(d) = d^{-2\alpha}$) is the most appropriate form for the kernel, and estimates that α is around 1.3 (Gibson 1997b). Figure 7.15, top graph, shows the instantaneous rate of infection (equal to the force of infection) for all susceptible trees at the start of 1982 using equation (7.19) with the proportionality constant derived to generate the observed number of new infections (45 cases) within one year. Due to the low value of the exponent in this transmission kernel, an infected tree could potentially infect many others outside the sampled orchard and, in fact, the average dispersal distance is infinite (see Section 7.4.2; Shaw 1995). Therefore, it is clear that the scale of an orchard can have an impact on the transmission potential from an infected tree. The bottom-left graph of Figure 7.15 shows how the total force of infection from a single infected tree in the center of a large square orchard scales with the size (number of trees) of the orchard, using the transmission kernel estimated for CTV. This result agrees with our hypothesis stated in Section 7.2.1.1—for plant diseases, R_0 will increase with the number of hosts.

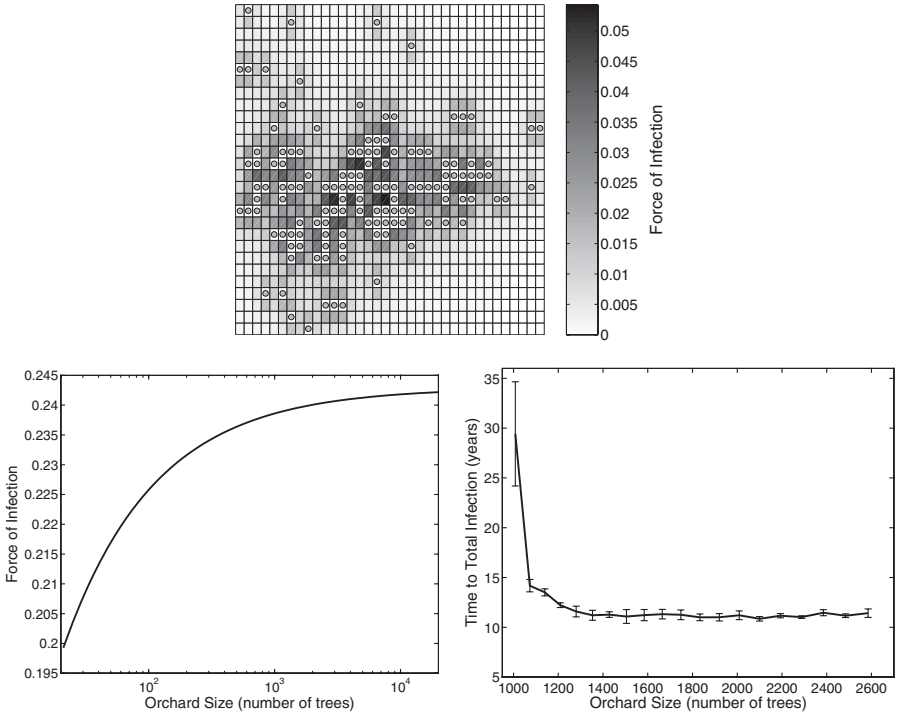


Figure 7.15. Graphs showing the transmission potential and dynamics of CTV, using the power-law kernel estimated by Gibson (1997b). The top graph shows the instantaneous force of infection to every susceptible tree at the end of 1981, with circles marking the position of infected trees. Note that the distance between trees is not equal; this is not a square lattice, there is greater spacing between the rows than the columns. The bottom-left graph shows the effect of orchard size on the force of infection from one infected tree in the center of the orchard. Finally, the bottom-right graph extends this force of infection result, by simulating the full epidemic dynamics across the whole orchard starting with the trees infected in 1981. The graph shows the average time for CTV to infect the entire orchard. The CTV data came from an orchard of size $28 \times 36 = 1{,}008$ trees; larger orchards were simulated by embedding the true orchard in the center of a larger grid (size $(28+n) \times (36+n)$ for n between 0 and 19).

The effect of orchard size is even more dramatic when the dynamics are iterated forward. Here orchard size plays two roles: (1) it increases the per capita force of infection, and (2) it increases the number of trees that will become infected in the next generation, hence speeding up the epidemic process still further. This secondary factor is promoted by the SI nature of infection such that trees remain permanently infected. The bottom-right graph of Figure 7.15 shows the average time (from the end of 1981) to infect an entire orchard of varying sizes. Despite the greater number of trees that need to be infected, larger orchards are predicted to be overwhelmed with infection quicker than smaller ones—although the time taken quickly saturates.

Although these simulations offer an understanding of CTV spread, several factors would need to be included if detailed predictions were required. The assumption of SI dynamics is clearly an approximation; infected trees do not instantaneously become infectious and infectious trees are eventually killed by the pathogen—therefore, $SEIR$-type dynamics may be a better assumption. The dynamics of the aphid vector have been ignored; the density of vectors is likely to vary throughout the year, producing some seasonal effects (see Chapter 5). Finally, the extrapolation from the small (28×36) sampled orchard to very large orchards assumes that the dispersal of the aphid vector is not influenced by the presence of trees; it is plausible to think that in a small orchard local transmission might be increased, because the aphids do not frequently leave the orchard, to regions with no fruit trees. However, despite these simplifying assumptions, the research by Gibson and coworkers have illustrated the difficulties in parameterizing transmission kernels from spatial data, but how sophisticated statistical techniques may provide a mechanism for teasing this information from the vast combinatorial array of possible infection scenarios.

7.5.2. Application: Spread of Foot-and-Mouth Disease in the United Kingdom

As a second example of individual-based modeling, we consider the model of Keeling et al. (2001b, 2003) for the spatial spread of foot-and-mouth disease in the United Kingdom. This model framework considers farms as the individual unit, classifying each farm as susceptible, exposed, infectious, or recovered. It is therefore assumed that once one animal in a farm is infected, the rest of the livestock on the farm soon contract the disease—this is a reasonable assumption given the highly transmissible nature of this virus. This model is far simpler than the general form of the individual-based model outlined above because farms do not move, and during the course of an epidemic farms will be neither created nor destroyed—except as part of the control measures. Therefore, the models of the foot-and-mouth epidemic are most closely related to those of plant epidemics, with the spatial dynamics being governed by a single transmission kernel.

The rate λ_i at which susceptible farm i becomes infected is given by:

$$\lambda_i = \mathrm{Sus}_i \sum_{j \in \text{infectious}} \mathrm{Trans}_j \, K(d_{ij}), \tag{7.20}$$

This is online program 7.6

$$\mathrm{Sus}_i = \sum_{l \in \text{species}} N_{i,l} \, s_l \qquad \mathrm{Trans}_j = \sum_{l \in \text{species}} N_{j,l} \, t_l,$$

where $N_{i,l}$ is the number of livestock of type l on farm i and s and t are species-specific susceptibility and transmissibility. Therefore, the farm-level susceptibility and transmissibility (Sus and $Trans$) are assumed to be the sum of the animal-level susceptibility and transmissibility for each species of livestock. The parameter values for s_{sheep}, s_{cattle}, t_{sheep}, and t_{cattle} were estimated by fitting to the spatial and temporal pattern of the 2001 UK epidemic, whereas the transmission kernel is derived from contact-tracing data that attempts to identify a source for each new reported case.

Figure 7.16 shows examples of the spatial and temporal dynamics of a simulated foot-and-mouth disease (FMD) outbreak in the United Kingdom if only minimal control measures are used (movement restrictions, which determine the kernel, and culling of animals on farms reporting infection). The spatial location of farms and the localized nature of the transmission kernel play a key role in determining the regions of the most severe outbreaks. However, despite the aggregation of cases, a huge prolonged epidemic

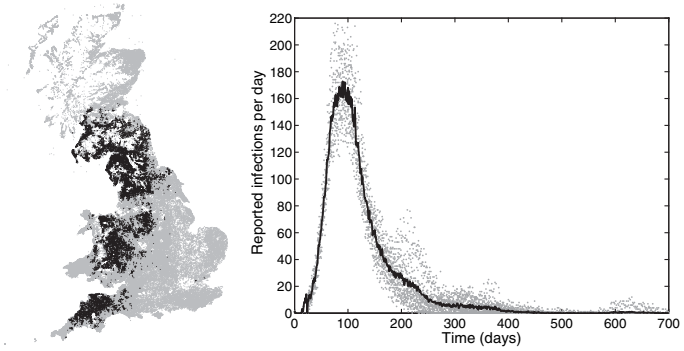


Figure 7.16. Results from the individual-based foot-and-mouth disease model, equation (7.20), using transmission parameters from Keeling et al. (2001b). Limited control measures are used—culling only animals on infected farms, together with movement restrictions—hence the epidemic is far larger, longer, and more widespread than in 2001. The left-hand figure shows the location of all livestock farms within the United Kingdom (gray dots) and those farms infected with FMD in a typical simulation (black dots). The right-hand graph gives the number of reported cases per day from ten simulations (gray dots) together with the average (black line). ($s_{sheep} = 1.0$, $t_{sheep} = 5.1 \times 10^{-7}$, $s_{cow} = 10.5$, $t_{cow} = 7.7 \times 10^{-7}$, $s_{pig} = t_{pig} = 0$, farms are latent for 5 days, then infectious for 5 days before they are culled). Simulation results kindly provided by M.J. Tildesley.

still ensues, lasting between one and two years and infecting more than 16,000 farms. The size of the outbreak is attributable to two main factors. First, foot-and-mouth disease has a relatively short latent and infectious period, and measuring the basic reproductive ratio as the number of secondary *farms* infected shows that $R_0 \approx 2.5$. This leads to a rapid doubling time of around a week in a totally susceptible population (Woolhouse et al. 2001a). In addition, animals (and therefore farms) are infectious for 4 to 5 days before clinic signs become apparent; therefore, a great deal of transmission may occur before it is realized that a farm has become infected (Fraser et al. 2004). This means that the removal of livestock from farms that are diagnosed with FMD is insufficient to control the disease.

In practice, additional reactive culling was performed as a preventative measure in an attempt to both remove farms that are likely to be infected and to remove susceptible farms that are likely to become infected in the future (Keeling et al. 2001b; Ferguson et al. 2001b). These culls took two main forms: Dangerous Contacts (DCs) found through the tracing of movements of vehicles, livestock, and people from an infected farm; and Contiguous Premises (CPs), defined as farms that share a common boundary with an infected farm. Unfortunately, neither of these culling practices is simple enough to be described in terms of a mechanistic formulation suitable for this book. Instead, we consider the likely success of ring culling (Ferguson et al. 2001a), removing the livestock on all farms within a given radius of an infected premise within 1 to 2 days. Figure 7.17 shows the impact of variously sized ring culls; although larger ring culls are always predicted to decrease the number of cases, the total number of farms lost (infected plus culls) is minimized for some intermediate culling policy of around 3.3 kilometers. Finally, the epidemic duration is minimized by large ring culls; therefore, a ring cull of 4 or 5 kilometers may be preferable if a long epidemic has severe economic or political consequences.

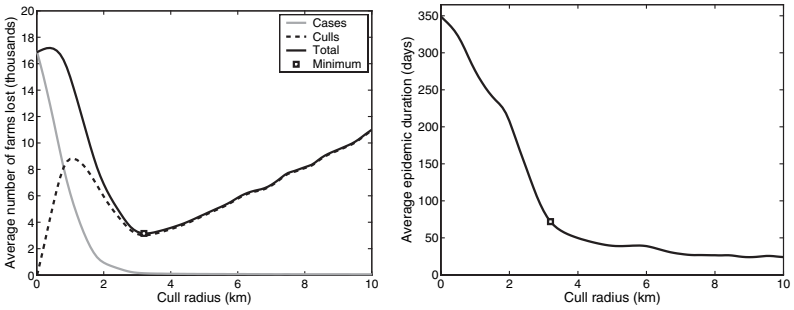


Figure 7.17. The simulated impact of ring-culling using the individual-based foot-and-mouth disease model, equation (7.18); with all other parameters as in Figure 7.16. The left-hand figure gives the average total loss of farms during an epidemic, composed of both reported infections and farms removed as part of a ring cull. There is a clear trade-off between minimizing the number of cases and not losing too many farms as part of the cull. For the same simulations the right-hand graph gives the average duration of the epidemic, with the cull size that minimized loss of farms marked. Results are from 1,250 simulations per radius value considered, and are smoothed using a local spline. Simulation results kindly provided by M. J. Tildesley.

7.6. NETWORKS

Networks provide a unified way to think about the interaction between individuals or populations, and are especially useful when each individual is in direct contact with only a small proportion of the population (Garnett and Anderson 1996; Morris 1997; Dunyak et al. 1998; Potterat et al. 1999; Klovdahl 2001; Rothenberg 2001; Sander et al. 2002; Potterat et al. 2002; Halloran et al. 2002; Liljeros et al. 2003; Rothenberg 2003; Szendroi and Csanyi 2004; Doherty et al. 2005; Keeling and Eames 2005). Networks tend to be very powerful tools for understanding the transmission of infection in human populations due to either social contacts (for airborne infections) or sexual contacts (for sexually transmitted diseases). In either case we expect that each individual will be in contact with only a small proportion of the population, and that the number of contacts will be highly heterogeneous—networks provide a simple means of capturing such interactions. We therefore see that the primary advantage of network models is their ability to capture complex individual-level structure in a simple framework.

To specify all the connections within a network, we can form a matrix from all the interaction strengths ρ_{ij}, which we expect to be sparse with the majority of values being zero. Usually, for simplicity, two individuals (or populations) are either assumed to be connected with a fixed interaction strength or unconnected (and therefore have an interaction strength of zero). In such cases, the network of contacts is specified by a graph matrix $\boldsymbol{G}$, where $\boldsymbol{G}_{ij}$ is 1 if individuals i and j are connected, or 0 otherwise. For the remainder of this section we will exclusively consider such networks where all the interaction strengths are identical; understanding networks with variable strength connections remains a challenge for the future (Sander et al. 2002). When $\boldsymbol{G}$ is symmetric ($\boldsymbol{G}_{ij} = \boldsymbol{G}_{ji}$) we define the network as undirected and infection can pass in both directions across a contact—this is the standard assumption for the vast majority of infectious diseases. However, there are a few special cases where a network is directed and infection

can pass only one way across a contact; examples of such directional transmission include infections transmitted through blood products, infections of livestock transmitted through artifical insemination, and transmission to populations (e.g., farms) through the movement of individuals (livestock).

Networks provide a robust means to consider the individual nature of disease transmission. Two individuals are linked if they have sufficient contact to allow the infection to pass between them.

A matrix G can be used to completely specify the network, indexing all possible transmission links between individuals.

7.6.1. Network Types

Several types of networks are commonly used within the epidemiological literature (as well as by statistical physicists). Although many theoretical approaches to networks use the terms nodes and edges, we will generally refer to individuals and contacts. Below we describe the basic nature of these networks in terms of a few fundamental properties: the way the network is constructed, the heterogeneity in the number of contacts, the clustering of contacts, and the average path length (in terms of the number of steps it takes on average to link two randomly chosen individuals). The five common network types discussed below are illustrated in Figure 7.18.

7.6.1.1. Random Networks

Random networks ignore the actual spatial position of individuals and, as the name suggests, connections are formed at random (Islam et al. 1996; Andersson 1998; Diekmann et al. 1998; Newman et al. 2002; Neal 2003). In one of the most analytically tractable versions of the random network, each individual has the same number of contacts. The random network is therefore characterized by a lack of heterogeneity in the number of contacts and a lack of clustering. The average path length in a random network is low because a large number of "long-distance" contacts exist, effectively spanning the population. A range of analytical techniques can be applied to understanding the dynamics of diseases spreading through such networks (Diekmann et al. 1998; Keeling 1999); of greatest importance, these models show that the initial growth rate of a disease in a network is reduced compared to the random-mixing equivalent:

$$\text{Initial random network growth rate} = \tau(\overline{n} - 2),$$

$$\text{Initial random-mixing growth rate} = \beta = \tau\overline{n},$$

where τ is the transmission rate across a contact and $\overline{n}$ is the average number of contacts (or effective number of contacts in the random-mixing model). This reduction is due to the development of strong negative correlations between susceptible and infected individuals during the early phase of the epidemic.

7.6.1.2. Lattices

As explained in Section 7.3, lattices are associated with a regular grid of contacts and each individual has a fixed number of contacts (usually either 4 or 8) (Bak et al. 1990;

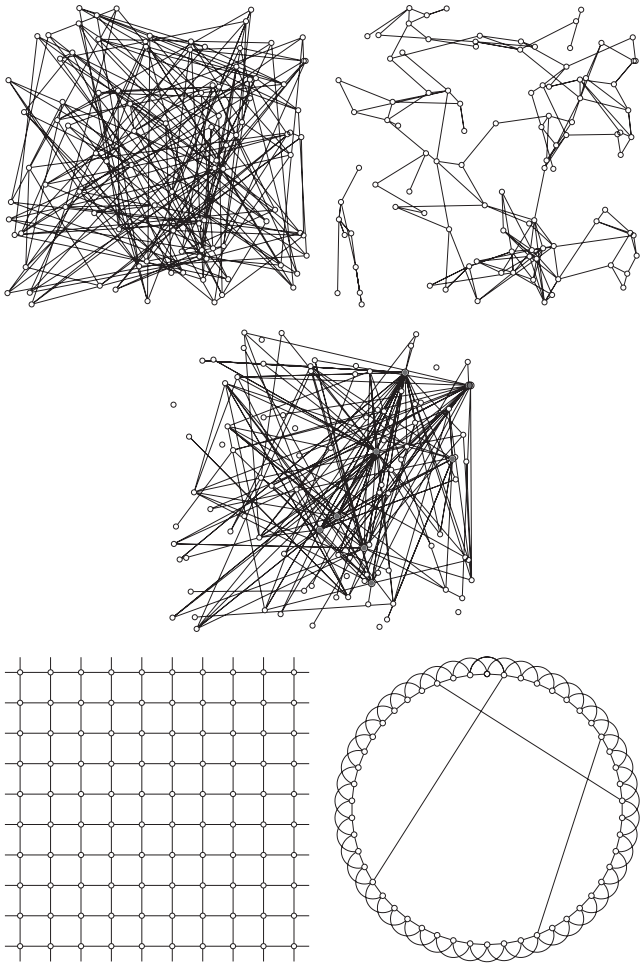


Figure 7.18. Five distinct network types containing 100 individuals. These are from left to right: Random, Spatial (top row), Scale-free (middle row), Lattice, and Small-World (bottom row). The Random, Spatial, and Scale-Free networks all use the same position of individuals—although for the Random and Scale-Free network, the position of the individuals is irrelevant for forming connections. In all five graphs, the average number of contacts per individual is approximately 4. For the scale-free network, individuals with high numbers of contacts are shaded gray.

Rhodes and Anderson 1997). In contrast to random networks, lattices possess far stronger clustering because contacts are localized in space. This higher level of clustering further reduces the initial growth rate of a disease in a lattice compared to the random network, although exact analytical results are no longer available. Given that all the connections are local, the average path length is very long because the only way to transverse the lattice from one side to the other is by steps of a single grid size.

7.6.1.3. Small World Networks

Small world networks are based upon a lattice structure, with a small number of "long-range" connections added. Figure 7.18 shows the classical one-dimensional small-world model (Watts and Strogatz 1998) where each individual is connected to its four nearest neighbors together with three long-range contacts across the entire population. Locally (from the perspective of an individual), small world networks look very much like lattices; they are highly clustered and have little heterogeneity in the number of neighbours—therefore, transmission of infection is predominantly localized so that the strong saturation effects and wavelike spread observed in the lattice models still occur. However, the presence of the few long-range connections provides shortcuts across the network, vastly reducing the average path length and allowing a spreading infection to jump to new susceptible areas (Newman and Watts 1999; Moore and Newman 2000). In practice, it may be difficult to estimate the number of long-range contacts, but small-world networks have highlighted their profound importance for disease dynamics (Boots and Sasaki 1999; Kuperman and Abramson 2001).

7.6.1.4. Spatial Networks

Spatial networks are one of the most flexible forms of networks, and are related to the individual-based models discussed in Section 7.5. In spatial networks, a kernel is often used to calculate the probability of any two individuals being connected depending on the distance between them (Watts 1999; Read and Keeling 2003; Keeling 2005a). By changing the distribution of individuals and the connection kernel, it is possible to generate a wide variety of networks from highly clustered lattices to small world arrangements and globally connected random networks. Spatial networks generally show a reasonably high degree of heterogeneity, with the number of neighbors often being approximately Poisson distributed. In addition, when the connection kernel preferentially links nearby individuals, we can regain the spatial wavelike spread of infection that characterizes lattice models.

7.6.1.5. Scale-Free Networks

In the vast majority of networks that have been studied, the number of contacts per individual is very heterogeneous, with most individuals having a relatively small number and a few have many contacts (Albert et al. 1999; Barabási and Albert 1999; Jeong et al. 2000; Lilijeros et al. 2001). Because the most connected individuals are likely to be disproportionately important in disease transmission (see Chapter 3), networks that can capture this heterogeneity are therefore vital in understanding the spread of real infections—scale-free networks incorporate these heterogeneities. Scale-free networks are generally created dynamically, adding new individuals to a network one at a time with a connection mechanism that mimics the natural formation of social contacts. Each new

individual that is added to the population connects preferentially with individuals that already have a large number of contacts; in a social setting, this corresponds to everyone wanting to be friends of the most popular people. The resultant network has a power-law distribution for the probability of having a given number of contacts, $\mathbb{P}(\text{contacts} > n) \approx n^{-\alpha}$. This power-law property was first observed for the World Wide Web connections and has also been recorded in power grid networks and graphs of actor collaborations. The same type of heterogeneities are likely to be present in the social contacts that permit the spread of infection.

Many different types of network structure are possible. These differ in the amount of heterogeneity, clustering, and average path length, thus reflecting the different transmission routes for various infections.

7.6.2. Simulation of Epidemics on Networks

Networks have many similarities with individual-based spatial models (see Section 7.5 and Section 6.3.5), in that spatial interactions can be defined in terms of a kernel. However, in networks, contacts tend to be of equal strength and limited in number. This can be used to considerable advantage in simulations:

$$\begin{aligned} Rate(\text{Infected individual } j \text{ recovers}) &= \gamma, \\ Rate(\text{Susceptible individual } i \text{ infected}) &= \tau \times \text{number of infectious contacts} \\ &= \tau \sum_j G_{ji} I_j = \lambda_i, \end{aligned}$$

where τ is the rate of transmission across a contact and I_j is one if individual j is infectious or zero otherwise. One immediate implication of the network structure is that the force of infection λ_i depends on the state of only a few individuals. This means that the force of infection does not need to be calculated anew at every iteration, providing huge computational savings. Instead, we can store the force of infection for each individual; when an individual first becomes infectious, the force of infection of all its contacts is increased by τ, and when an individual recovers the force of infection for all its contacts is likewise increased by τ. Hence, each event impinges on the state of only its neighborhood of contacts. Further computational savings can be achieved if the contacts of each individual are stored in a list (such that $C_1^i, C_2^i, \ldots C_{n_i}^i$ are the n_i contacts of individual i) because loops and summations need to be over only the n_i neighbors, which is generally much faster than summing over the total number of individuals. For example:

$$Rate(\text{Susceptible individual } i \text{ infected}) = \tau \sum_{j=1}^{n_i} I_{C_j^i}.$$

This is online program 7.7

Figure 7.19 shows examples of the epidemic dynamics on the network types illustrated in Figure 7.18, giving both the individual epidemic curves (gray lines) and the average of all major epidemics (black line). Clearly the random network (which in many ways is closest to the nonspatial mass-action models) generates the fastest epidemic growth rate and has the highest proportion of infectious individuals at the maximum. Surprisingly, there appears to be little difference between the dynamics of epidemics on the Spatial and Scale-Free networks; this may be attributable to the variance in the number of contacts that is present in both networks. For infinitely large population sizes, Scale-Free

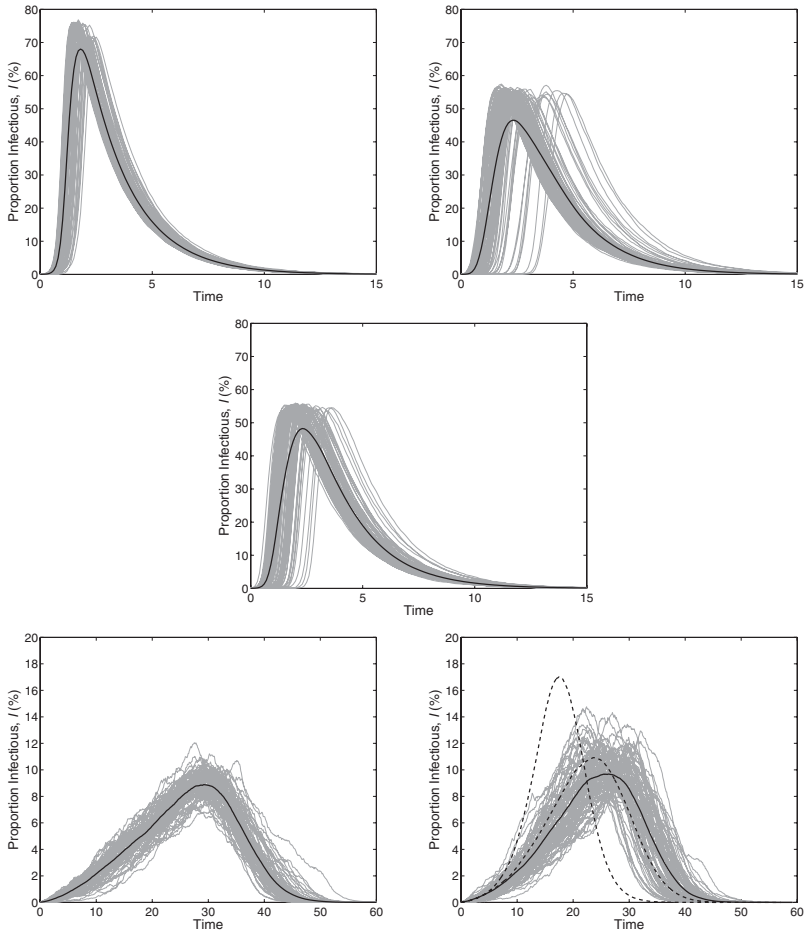


Figure 7.19. Typical epidemics of the five network types described above. These are from left to right: Random, Spatial (top row), Scale-Free (middle row), Lattice, and Small-World (bottom row). Each graph shows 100 epidemic curves (gray), together with the average for all major epidemics (black) for a single example of each network type—therefore, all variablity within each graph is due to the stochastic nature of transmission and not variation in the network. All five networks contain 10,000 individuals, although all individuals are not necessarily interconnected as part of a giant component. For the Spatial and Scale-Free networks, approximately 88% and 74% are part of the giant component and can therefore potentially become infected; for these networks, the proportion of infectious individuals have been rescaled as a fraction of the giant component. For all other types, the entire network is interconnected. In all networks, the average number of contacts per individual is approximately 4, although the Scale-Free network has considerable heterogeneity, with one individual having 85 contacts. For consistency, the Small-World network is formed from a two-dimensional lattice (Watts 1999) (not a one-dimensional circle as shown in Figure 7.18) with 10 additional random "long-range" contacts. The dashed lines show the effect on the mean epidemic of increasing the number of long-range contacts to 20 and 100. ($\tau = 1$, $\gamma = 0.5$.)

networks theoretically have $R_0 = \infty$ because there will be some individuals with arbitrarily large numbers of contacts (Albert et al. 2000; May and Lloyd 2001). However, in any practical scenario, the maximum number of contacts is always limited and the success of the epidemic depends on encountering these core individuals early.

In contrast, the lattice and two-dimensional small-world networks show much slower epidemic growth rates. In fact, as predicted in Section 7.3, the lattice network leads to an expanding wave of infection and hence an (almost) linear initial increase in the number of cases. As long-range contacts are added to this basic lattice, forming a Small-World network (Watts and Strogatz 1998), the infection is able to spread to new susceptible areas of the network and therefore the infection grows more rapidly.

In general, networks display slower epidemic dynamics compared to randomly mixed models. As a consequence, networks that are most like the random-mixing models—with short-average path length (Small-World, Random, and Scale-free) and little clustering (Random, and Scale-Free)—show the fastest epidemic growth rates for a given average number of contacts per individual.

7.7. WHICH MODEL TO USE?

With such a vast array of possible spatial models, it may be daunting to try to choose between them. As with all model choices, the type of model required will reflect the problem being addressed, the availability of data, and the form of results required. At some fundamental level, *all* spatial models can ultimately be expressed as both metapopulation models or individual-based models with carefully constructed interaction terms. However, the following guide may help discriminate between the model classes:

- *All individuals interact at random.* A spatial model is not required.
- *The population is naturally separable into groups, with strong (random) interaction within each group.* This is the classical metapopulation ideal, and is frequently the preferred model for human disease dynamics at a national scale where the population can be generally grouped by town or city.
- *The population is densely distributed across the entire space.* Here we may safely treat the population as continuous and deterministic; therefore, either PDE or integro-differential models can be appropriate.
- *The environment and distribution of hosts is approximately uniform and a qualitative understanding of spatial effects is required.* In such situations, the approximate lattice-based models may be suitable. It should be noted, however, that such models are unlikely to provide an accurate prediction of the quantitative behavior of any real problem.
- *Hosts have a low density or patchy distribution and stochasticity effects are important.* Here we must resort to individual-based modeling; this has the extra advantage that greater behavioral complexity can be easily included. Note that parameters are generally defined at the individual level, such that aggregate population-level data is difficult to use.
- *Hosts have few contacts to whom they can pass infection.* In such cases, networks are the preferred modeling tool. Again, this is an individual-based approach, such that individual-level data is required for parameterization.

Although the above points provide a general guide to the types of models that can be used, compromises are frequently involved. Often individual-based models will seem to be the preferred model type because they are the most flexible and can most closely mimic reality. However, in a great many scenarios there is insufficient information to parameterized such models—for example, only aggregate data may be available. Individual-based models can also be extremely slow compared to other techniques and it may be difficult to assess the robustness of the model to the wealth of factors that can be included. We therefore see that model choice is a skill in itself; ideally several model formats should be tested, their output compared to the available data, and the predictions of all the models scrutinized in terms of the different elements that have been ignored. The variety of models used during the 2001 foot-and-mouth disease outbreak in the United Kingdom illustrate the advantages of this approach where widely different model assumptions all produced similar control recommendations (Keeling et al. 2001b; Ferguson et al. 2001a,b; Morris et al. 2001; Keeling 2005b). However, the use of multiple models to investigate the use of vaccination in the case of a smallpox outbreak produced conflicting recommendations (Meltzer et al. 2001; Kaplan et al. 2002; Halloran et al. 2002; Bozzette et al. 2003)—highlighting the sensitivity of this problem to model structure and assumptions (Ferguson et al. 2003b). Obviously, such a comprehensive approach is rarely possible for a single researcher (or research team); instead, the merits and assumptions of each approach must be weighed against the available data and the required detail and accuracy of model predictions.

7.8. APPROXIMATIONS

In general, spatial models tend to be computationally intensive, such that most results can be considered as in silico experiments, rather than definitive answers. Researchers are therefore beginning to consider other means of modeling spatial epidemics, such that some of the robustness and understanding that comes from the standard differential equation models can be regained. This approach can be compared to the various approximations that have been used to understand stochastic systems (Chapter 6). Here we briefly consider two forms of approximation, both based on the idea of modeling pairs of hosts and hence capturing the correlations that develop when two individuals interact.

Although more mathematically involved than the standard simulation models described in Sections 7.2 to 7.6, these pair-wise models offer the chance to understand the processes involved in disease transmission in a spatial environment. In both approximation approaches given below, negative correlations between susceptible and infectious individuals are of paramount importance for the dynamic behavior, reducing the growth of epidemics by effectively reducing the transmission. This is a universal feature that we have observed in all the spatial models of this chapter; however, it is only through the use of analytical approximation methods (such as those illustrated below) that these effects become fully apparent.

7.8.1. Pair-Wise Models for Networks

The simplest form of a pair-based approximation model is used to capture disease spread through a network of contacts (see Section 7.6). In its most basic form, this pair-wise model assumes an equal number of contacts per individual and no clustering; this

approximation therefore corresponds most closely to the random network (Keeling et al. 1997b; Keeling 1999; Bauch and Rand 2000), although adaptations that capture clustering or heterogeneities are possible (Keeling 1999; Eames and Keeling 2002; Eames and Keeling 2004). To explain the formulation of pair-wise models, we go back to the original SIR (or SIS) models discussed in Chapter 2:

$$\text{Rate of new infection} = \text{transmission rate} \times \text{number of susceptibles} \times \text{number of contacts} \times \text{probability contact is infectious.}$$

In the standard random-mixing (frequency-dependent) models this was approximated by:

$$\text{Rate of new infection} \approx \tau \times X \times n \times Y/N = \beta XY/N,$$

where τ is the transmission rate between contacts, n is the average number of contacts, and Y/N is an approximation for the probability that a contact is infectious. This latter term is an approximation because it neglects all spatial correlations between connected susceptible and infectious individuals.

If we know the number of susceptible-infected pairs (i.e., the number of susceptible individuals in contact with an infectious individual in the network), which we label $[XY]$, then the calculation of the infection dynamics is exact:

$$\text{Rate of new infection} = \tau[XY].$$

This calculation is exact because $[XY]$ takes into account all local spatial correlations. (We can think of the mean-field models as approximating $[XY]$ by nXY/N.) However, if we wish to use this pair-wise technique predictively, we need to model how $[XY]$ varies over time. We therefore develop a differential equation for the dynamics of $[XY]$ pairs. For the SIR equations:

$$\frac{d[XY]}{dt} = \tau[\overleftarrow{XX}Y] + \gamma[YY] - \tau[\overleftarrow{XY}] - \tau[\overrightarrow{Y}\overleftarrow{XY}] - \gamma[XY]. \tag{7.21}$$

Here an arrow signifies the direction of transmission and triples ($[ABC]$ represents an A connected to a B connected to a C in the network) are spaced such that the pair in question is more clearly identified. Five events can lead to changes in the number of XY pairs; in the order they appear in equation (7.21) these are: creation of an XY pair by infection of an XX pair, creation of an XY pair by recovery of an infected individual in a YY pair, loss of an XY pair due to the susceptible being infected by the infectious individual within the pair, loss of an XY pair due to the susceptible being infected by an infectious individual outside the pair, and loss of an XY pair due to recovery of the infected individual.

To close the dynamics, we need to know the number of triples—in particular [XXY] and [YXY]. We could formulate an equation for the number of triples, which would contain expressions involving the number of quads. However, it is simplest to perform a moment-closure approximation, by approximating the number of triples in terms of the number of pairs and singles. If all individuals within the contact network have exactly n contacts, then the triple approximation becomes:

$$[ABC] \approx \frac{(n-1)}{n}\frac{[AB][BC]}{[B]}.$$

This approximation therefore ignores any correlation that may have developed between the ends of the triple—that is, A and C are correlated only by the fact that they are both connected to B. Therefore, if triples also form triangles, such that A and C are connected

in the network, then this approximation is likely to be flawed, although still better than the mean-field approximation that ignores all correlations.

Box 7.4 SIS Pair-Wise Equations

Although equation (7.22) provides the most intuitive description of the pair-wise network approximation for the SIS model, it is informative to show the full set of equations, where the triple approximation has been included and the dynamics are written in terms of just two state variables:

$$\frac{dX}{dt} = \gamma(N-X) - \tau[XY],$$

$$\frac{d[XY]}{dt} = \tau\frac{(n-1)}{n}\frac{(nX-[XY])[XY]}{X} + \gamma(nN-nX-[XY]) - \tau[XY] - \tau\frac{(n-1)}{n}\frac{[XY]^2}{X} - \gamma[XY].$$

We are now in a position to formulate a pair-wise equation for the dynamics of a disease on a random (unclustered) network where each individual has exactly n contacts. For a disease with SIS dynamics and no births or deaths:

$$\begin{aligned}\frac{dX}{dt} &= \gamma Y - \tau[\overleftarrow{XY}] \\ \frac{d[XY]}{dt} &= \tau[\overleftarrow{XXY}] + \gamma[YY] - \tau[\overleftarrow{XY}] - \tau[\overrightarrow{YXY}] - \gamma[XY].\end{aligned} \tag{7.22}$$

This is online program 7.8

Noting that

$$[ABC] \approx \frac{(n-1)}{n}\frac{[AB][BC]}{[B]}, \qquad Y = N - X, \qquad [XX] = nX - [XY].$$

Figure 7.20 shows a comparison between stochastic SIS epidemics on a random network and results from the deterministic pair-wise model. Clearly there is excellent agreement between the two approaches, even though the pair-wise model is deterministic and requires only two equations, whereas the network simulation is stochastic. However, the advantage of the pair-wise model is in its analytical tractability and its comparison to the standard differential equations (Chapter 2) for disease dynamics.

Network-based pair-wise models provide a deterministic approximation for the dynamics of pairs of connected individuals within a network, such as the number of connected susceptible-infectious pairs, and can therefore account for the buildup of local correlations within the network.

Network-based pair-wise models are most accurate when dealing with networks with low levels of clustering, such as Random or Scale-Free networks.

Network-based pair-wise models approximate the buildup of local correlations within the network by explicitly modeling connected pairs of individuals.

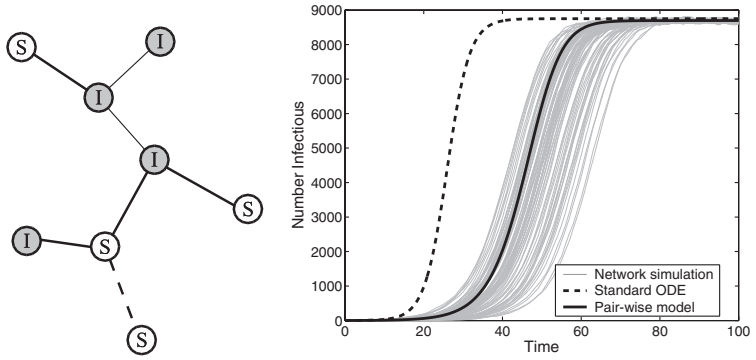


Figure 7.20. The left-hand figure shows a caricature of a network, so that pairs of connected individuals can be readily seen. This particular example has four $[SI]$ pairs (and hence also four $[IS]$ pairs), indicated with thick black contacts; four $[II]$ pairs, because these can be counted in both directions, shown as thin contact lines; and two $[SS]$ pairs, again counting in both directions. The right-hand graph compares the results for a stochastic SIS epidemic on a Random network (see Section 7.5, Figure 7.19) with the corresponding pair-wise equation (7.22) and the standard ODE model (see Chapter 2). Clearly, the correlations in the pair-wise model capture the reduced epidemic growth rate seen in the full network model. ($N = 10,000$, $n = 4$, $\tau = 0.1$, $\gamma = 0.05$.)

7.8.2. Pair-Wise Models for Spatial Processes

A similar approach can be used for approximating the dynamics of full individual-based spatial models (see Section 7.5). In the network-based pair-wise models, a contact was either present or absent, so that all XY pairs had the same strength of transmission. However, for individual-based models, the transmission strength is a function of distance, determined by a transmission kernel; therefore, all XY pairs must be indexed by the distance between them. We therefore write $[XY](\underline{d})$ as the density of XY pairs separated by a vector distance $\underline{d}$; however, in many situations it is easier to consider this pair-wise quantity to be composed of the mean densities plus the spatial covariance (the covariance between two distinct spatial points):

$$[XY](\underline{d}) = X \times Y + C_{XY}(\underline{d}),$$

where X and Y can now be thought of as densities, or more formally probability densities, for an individual to exist at a particular point in space. For such models, the calculation of the rates of change is somewhat more involved due to the nature of individual-based spatial transmission:

$$\begin{aligned}\text{Rate of new infection} &= \tau \int K(\underline{r})[XY](\underline{r})d\underline{r}, \\ &= \beta XY + \tau \int K(\underline{r})C_{XY}(\underline{r})d\underline{r},\end{aligned}$$

where K is again the transmission kernel that measures how infection risk decreases with distance, and β $(= \tau \int K(\underline{r})d\underline{r})$ is again an aggregate measure of transmission. For the rate

of change of pairs we have:

$$\begin{aligned}\frac{d[XY](\underline{d})}{dt} &= \tau[XX](\underline{d})\frac{\int K(\underline{r})[XY](\underline{r})[XY](\underline{r}-\underline{d})d\underline{r}}{X^2Y},\\ &\quad -\tau[XY](\underline{d})\frac{\int K(\underline{r})[XY](\underline{r})[YY](\underline{r}-\underline{d})d\underline{r}}{XY^2},\\ &\quad -\tau[XY](\underline{d})K(\underline{d})+\gamma[YY](\underline{d})-\gamma[XY](\underline{d}).\end{aligned} \tag{7.23}$$

Here the first term corresponds to the infection of one susceptible (of a XX pair separated by a distance $\underline{d}$) by an infectious individual who is distance $\underline{r}$ from the target susceptible and therefore distance $\underline{r}-\underline{d}$ from the other susceptible within the pair. This term is composed of three pair-wise components, because all three individuals involved in the event have associated correlations—this differs from the network approach where all triples are "linear"; here all triples form triangles. The second term refers to the loss of an XY pair due to the susceptible being infected from outside the pair. The other three terms do not involve any external interactions and are: infection within an XY pair, recovery within a YY pair, and recovery within an XY pair.

Pair-wise models for spatial processes provide a deterministic approximation for individual-based spatial models.

Although the solution of complex integro-differential equations such as equation (7.23) is an involved process, for some spatial kernels semi-analytic results are possible (Bolker and Pacala 1997; Bolker 1999; Dieckmann et al. 2000) and hence a more comprehensive and robust understanding of the general effects of spatial interaction are possible.

7.9. FUTURE DIRECTIONS

It has only been over the past decade that spatial models have been used in epidemiological applications, and in many areas the implications of spatial structure have yet to be understood. Two main problems stand out as being of both theoretical and applied importance.

The first is the theoretical question of scale—which has applied implications in terms of what is computationally feasible. It is clear from the work in this chapter that spatial structure occurs and operates at a variety of scales; however, it is not clear whether it is necessary to simulate all of these scales to derive accurate predictions of the epidemiological dynamics. In particular, is it always necessary to model the underlying network structure of social contacts to create reliable epidemic models at the national (or international) level? This is one aspect of a more general issue: When can complex fine-scale heterogeneities be absorbed into parameterization and when must they be modeled explicitly? The approximation methods outlined above may be able to provide some insights and lead to general rules about when spatial structure is important.

Second, many difficulties still exist with parameterizing spatial models, either from epidemiological or behavioral data. This is exemplified by three different problems. First, we have very little information about the network of social contacts through which most airborne infections spread, although this has recently become an area of intense focus. Understanding the role of transmission network structure is fundamental to all infectious

disease modeling, and may validate or refute the standard (pseudo) mass-action assumption in particular circumstances. Second, at a larger scale, the 2001 foot-and-mouth epidemic in the United Kingdom highlighted the difficulties with assessing the degree of spread between farms. In reality, this spatial spread is the combination of several factors (such as the movement of vehicles or wind-borne transmission), but due to the complexities of parameterization this spread was estimated and modeled as a single transmission kernel. Finally, metapopulation models—treating each town or city as a subpopulation—appear to be an ideal tool for modeling national epidemic patterns; however, it is difficult to assess the level of coupling between communities, especially when many of the links may be sporadic and social.

Spatial models will continue to be an area of high research activity for many years. The importance of local spatial interaction is only recently being appreciated, in terms of both understanding disease dynamics and local control of infection. We can expect to see spatial models increasingly used in public health scenarios, where control usually operates on a regional basis and where preventing infection reaching new populations is a key control aim.

7.10. SUMMARY

Seven different model formulations have been described in this chapter (Metapopulations, Coupled lattices, Cellular automata, Reaction-Diffusion, Integro-Differential, Individual-Based and Network), each with its own merits and disadvantages. The techniques needed to simulate these models and the data needed to parameterize them differ greatly; however, they share a common theme in that local interactions (transmission) generally dominate longer-range interactions, leading to a clustering of cases.

➤ The type of spatial model used is dependent on the host organism, our degree of knowledge about its behavior, and the scale we wish to consider.

➤ **Metapopulations** provide a powerful framework for modeling disease dynamics for hosts that can be naturally partitioned into spatial subunits.

➤ The force of infection within a subpopulation can be modeled as a weighted sum of the prevalence in all subpopulations.

➤ With stochastic metapopulations, the speed of the spread of infection between subpopulations is reduced compared to the equivalent deterministic model.

➤ For plants and other sessile hosts, coupling (or the strength of spatial interaction) generally decreases with distance, mimicking the effects of wind—or vector—dispersal. Adding an extra subpopulation generally increases R_0 because more pathogens can be intercepted by the additional hosts.

➤ Metapopulation models of animal diseases usually capture the transmission of infection by the permanent immigration and emigration of hosts. In these models, R_0 is generally independent of the coupling because each host transmits infection at a constant rate.

➤ The spread of human diseases in metapopulations is best captured by the rapid commuter movements of individuals from their home subpopulation to another subpopulation and back again—requiring us to model both the current location and home location of

individuals. When commuter movements are of short duration, this can be approximated by simple coupling. In these models, R_0 is independent of the coupling.

➤ The correlation between disease prevalence in two subpopulations increases sigmoidally with the strength of interaction between the populations. In general, the change from largely independent dynamics to synchrony occurs for interaction strengths from 10^{-3} to 0.1.

➤ Without interaction between the subpopulations, a spatially segregated metapopulation suffers a faster rate of stochastic extinction than its randomly mixed counter-part. When interaction between the subpopulations is included, the level of local (subpopulation-scale) and global (metapopulation-wide) extinctions is an emergent property of the dynamics and cannot be easily predicted from the parameters.

➤ **Levins' metapopulation** models ignore the internal dynamics within each subpopulation, and instead classify each subpopulation as either infected or disease free. Despite differences between the equilibrium-level results of Levins and full metapopulation models, the Levins model still remains a useful and simple tool for studying invasion dynamics.

➤ **Coupled lattice models** are specialized metapopulation models, where subpopulation are arranged on a grid and coupling is generally to the nearest neighbors only.

➤ The wave speed of an invading epidemic in a coupled-lattice model increases almost linearly with the initial growth rate of the infection, $\beta - \gamma - \nu$; increases nonlinearly with the level of coupling, ρ; and is slightly more rapid in deterministic compared to stochastic models.

➤ **Cellular automata** operate on a lattice of sites, with each site generally assumed to hold a single host. Interaction is usually stochastic and with the neighboring (four or eight) lattice sites.

➤ In many locally coupled spatial models (such as cellular automata), the depletion of the locally available susceptible population can reduce the early growth rate of the epidemic and the speed of the invading wave front.

➤ The forest-fire model typifies many stochastic spatial cellular automata models. The fact that transmission is faster than recovery, which is faster than births, which is faster than imports of infection, leads to power-law relationships for the frequency of epidemic sizes.

➤ **Reaction-diffusion models**, which use a PDE formulism, assume local transmission of infection and rely on spatial diffusion of hosts to spread the infection.

➤ For PDE models (in two dimensions with equal diffusion in all directions), infection spreads as a growing circular wave of near constant velocity.

➤ With **integro-differential equations**, the spatial spread of infection is via a transmission kernel that defines how transmission risk decays with distance. The shape of the tail of the transmission kernel determines the eventual spatial pattern of invasion, from wavelike spread, to scattered local foci, to highly probable extremely long-range jumps.

➤ **Individual-based models** account for the spatial interaction between individual hosts distributed on a spatial landscape. They can include a wide variety of complex (more

biologically realistic) behavior that often features a spatial component; this can lead to a huge number of parameters that can be difficult to determine from available data.

➤ **Networks** provide a robust means to consider the individual nature of disease transmission. Two individuals are linked if they have sufficient contact to allow the infection to pass between them.

➤ Many different types of network structure are possible. These differ in the amount of heterogeneity, clustering, and average path length, thus reflecting the different transmission routes for various infections.

➤ In general, networks display slower epidemic dynamics compared to randomly mixed models. As a consequence, networks that are most like the random-mixing models—with short average-path length (Small-World, Random, and Scale-Free) and little clustering (Random and Scale-Free)—show the fastest epidemic growth rates for a given average number of contacts per individual.