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:

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

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:

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

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.

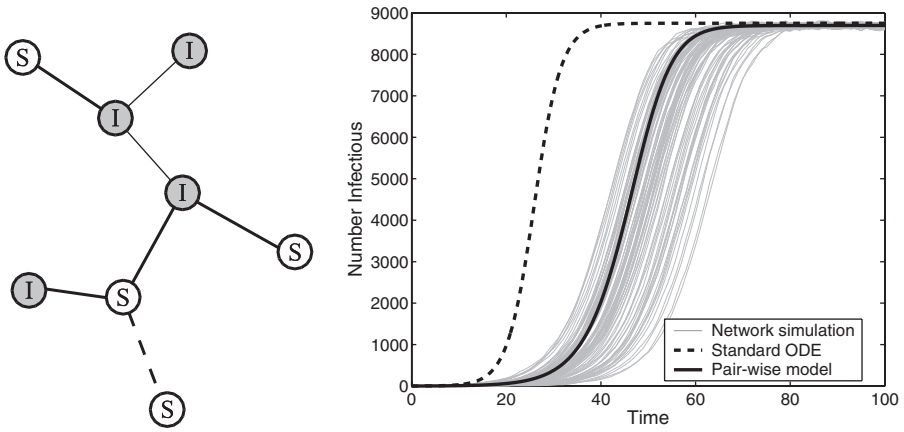

Figure 7.20. The left-hand figure shows a caricature of a network, so that pairs of connected individuals can be readily seen. This particular example has four $[SI]$ pairs (and hence also four $[IS]$ pairs), indicated with thick black contacts; four $[II]$ pairs, because these can be counted in both 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