Correlation models for childhood epidemics

M. J. KEELING¹, D. A. RAND² AND A. J. MORRIS²

¹*Department of Zoology, Downing Street, University of Cambridge, Cambridge CB2 3EJ, UK*

²*Nonlinear Systems Laboratory, Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK*

SUMMARY

One of the simplest set of equations for the description of epidemics (the SEIR equations) has been much studied, and produces reasonable approximations to the dynamics of communicable disease. However, it has long been recognized that spatial and social structure are important if we are to understand the long-term persistence and detailed behaviour of disease. We will introduce three pair models which attempt to capture the underlying heterogeneous structure by studying the connections and correlations between individuals. Although modelling the correlations necessarily leads to more complex equations, this pair formulation naturally incorporates the local dynamical behaviour generating more realistic persistence. In common with other studies on childhood diseases we will focus our attention on measles, for which the case returns are particularly well documented and long running.

1. INTRODUCTION

Measles is one of the most infectious communicable diseases. Prior to immunization, measles was a common childhood disease even in developed countries with over 90% of the population having been infected by age 20. In young and malnourished children measles is fatal in between 5% and 10% of cases and so is responsible for many deaths in the developing world (Benenson 1990).

The prediction of measles epidemics and dynamics is undoubtedly very important from a public health view. On the other hand, the understanding of measles has become a goal in itself because, as well as possessing complex dynamics, it has a very simple natural history and therefore many plausible population models (Hamer 1906; Bartlett 1957; Schwartz 1985; Olsen 1987; Schenzle 1984; Bolker & Grenfell 1993). The data for measles are very long and detailed in comparison to other ecological time series. It is therefore possible to obtain a picture of the long-term dynamics of this disease. It is periodically forced by the school year, and although biennial cycles predominate, annual and three-year cycles are also observed (Fine & Clarkson 1982; Schenzle 1984). One of the main stumbling blocks to a more complete understanding of the disease is the inability of models to match the existence of relatively violent seasonally driven epidemics with persistence at the critical population size of around 300 000–500 000 (Bartlett 1960; Black 1966).

After first reviewing the simple SEIR model, we develop a correlation model which considers the behaviour of connected pairs. Because it takes account of the rapid build up of strong correlations, this model does not suffer from the deep troughs associated with the SEIR model, and can produce oscil-

latory dynamics without resorting to seasonal forcing (cf. Johansen 1996). Based on this pair formulation two further models will be introduced. The first is a simplification which illustrates the main differences between the mean-field and pairwise approaches, the second is a more complex age-structured model which produces convincing dynamics, both biennial cycles and irregular two- to three-year epidemics. The age-structured model generates a far more realistic fade-out pattern than its mean-field counterpart, and shows power-law scaling of epidemics—a feature usually associated with full spatial models (Rhodes & Anderson 1996*a,b*). In contrast to this previous work, we predict that in general there should be distinct regions of different power-law behaviour reflecting different scales of population structure. For measles in developed countries, we predict three regions corresponding to within-family, within-school and between-school transmission. Thus the scaling structure gives insight into crucial aspects of population structure.

2. MEAN-FIELD DYNAMICS

Due to life-long immunity to reinfection, measles epidemics are self-extinguishing as waves of infection. However, new susceptible individuals are born behind these waves and this, together with the periodic forcing of the school year, produces complex dynamics. The well-mixed situation can be fairly well described by the SEIR equations,

$$\left. \begin{aligned} \dot{S} &= m(E + I + R) - \beta SI, \\ \dot{E} &= \beta SI - (m + a)E, \\ \dot{I} &= aE - (m + g)I, \\ \dot{R} &= gI - mR, \end{aligned} \right\} \quad (1)$$

Figure 1. The number of reported cases of measles in England and Wales over the prevaccination period from 1948 to 1966. Strong biennial cycles dominate the dynamics.

which model the proportions of susceptible, exposed, infectious and resistant individuals in the population (Anderson & May 1992). In these equations the mortality ($m \approx (50 \text{ yr})^{-1}$) is equal to the per capita birth rate and both are independent of the previous history of the individual. a^{-1} is the latent period of the infection (8 days), while g^{-1} is the infectious period (5 days) (values from Anderson & May 1992). β is the effective contact and transmission rate, it is the number of the susceptibles contacted by a single infectious individual that catch the disease assuming total homogeneous mixing. β is often modelled as seasonally varying to mimic the greater amount of contact during the school terms,

$$\beta = 4.93(1 + 0.28 \cos(2\pi t)) \text{ day}^{-1},$$

where t is the time in years (values from Olsen & Schaffer (1990)). With this amount of seasonal variation, the SEIR equations possess chaotic dynamics and produce large irregular epidemics. The application of methods from nonlinear dynamical systems to equation (1) and to the time series (figure 1) has led to considerable progress in the study of measles dynamics (Aron & Schwartz 1984; Rand & Wilson 1991; Grenfell *et al.* 1995). Although equation (1) captures the large erratic peaks of infection, there are long periods of time when the disease has effectively died out as the number of infectious individuals would be far less than one for any reasonable population. Moreover, the actual data from England and Wales has fairly clear self-sustained cycles of about two years, although data from some other countries exhibit occasional periods of three years. The SEIR equations, however, show no regular periodicity and this together with its frequent extinctions necessitates the use of more complex models.

3. A PAIR CORRELATION METHOD

In previous work, predefined spatial correlations have been mainly used to model the aggregation of particular individuals (e.g. Hassell & May 1974);

however, in many explicit spatial models and real ecologies strong dynamical correlations exist between different species. We wish to model these.

We will consider the modelling of a network of individuals or sites whose states are changing at rates determined by the state of their neighbours. This is analogous to a probabilistic cellular automaton but we allow far greater flexibility in the statistical structure of the neighbouring sites. The standard mean-field equations (1), that are used as the basic approximation to the dynamics of such a system, assume that the states in the neighbourhood are independent of the state of the centre site, i.e. there are no correlations. As an improved approximation, equations for the behaviour of pair correlations can be derived from the basic process (Matsuda *et al.* 1992; Satō *et al.* 1994). A pair is considered to be any two sites where direct transmission of the disease between the pair is possible. This is an implicit spatial model. The fact that a system is distributed over space leads to local correlations and by describing these the spatial nature of the system is implicitly modelled.

We shall adopt the following notation: let $[X]$ be the number of sites that are of type X , let $[XY]$ be the number of X - Y pairs and similarly $[XYZ]$ be the number of X - Y - Z triples. For simplicity of notation we shall adopt the following counting convention; X - Y pairs are counted once in each direction ($[XY] = [YX]$) and X - X pairs are counted twice ($[XX]$ is even). It should be noted that the number of singles can be calculated from the number of pairs:

$$[X] = \frac{1}{n} \sum_Y [XY],$$

where n is the number of neighbours of a site, hence it is only necessary to reformulate equation (1) in terms of pairs. The pair equations are deduced from the fact that the rate of change of the expectation of $f = [AB]$ is the sum over all possible events of the rate of the event multiplied by the change the event causes in f .

$$\left. \begin{aligned} \dot{[S]} &= 2m([SE] + [SI] + [SR]) - 2\tau[SSI], \\ \dot{[SE]} &= m([EE] + [EI] + [ER] - [SE]) \\ &\quad + \tau([SSI] - [ESI]) - a[SE], \\ \dot{[SI]} &= m([EI] + [II] + [IR] - [SI]) \\ &\quad - \tau([ISI] + [SI]) + a[SE] - g[SI], \\ \dot{[SR]} &= m([ER] + [IR] + [RR] - [SR]) \\ &\quad - \tau[RSI] + g[SI], \\ \dot{[EE]} &= -2m[EE] + 2\tau[ESI] - 2a[EE], \\ \dot{[EI]} &= -2m[EI] + \tau([ISI] + [SI]) \\ &\quad + a([EE] - [EI]) - g[EI], \\ \dot{[ER]} &= -2m[ER] + \tau[RSI] - a[ER] + g[EI], \\ \dot{[II]} &= -2m[II] + 2a[EI] - 2g[II], \\ \dot{[IR]} &= -2m[IR] + a[ER] + g([II] - [IR]). \end{aligned} \right\} \quad (2)$$

To close the above equations it is necessary to approximate the various triples that appear. It should

be noted that triples only occur in terms of the transmission of infection as only then is the neighbourhood of a pair important. The most obvious assumption would be to suppose that all triples are formed from two pairs, which share a common central individual but are otherwise independent,

$$[XYZ] \approx \frac{(n-1)[XY][YZ]}{\sum_W [YW]} = \frac{n-1}{n} \frac{[XY][YZ]}{[Y]}. \quad (3)$$

However, this naively removes a great deal of spatial structure which may be inherent in the topology of the contact network. This can clearly be seen if there is a significant number of triangular paths present in the network (see figure 2), in this case an X - Y - Z triangular triple also contains the pair $[XZ]$ and this will inevitably have an effect. We shall let ϕ be the proportion of triples that form triangles. To handle this situation, Morris (1997) and van Baalen (personal communication) have proposed the following approximation:

$$[XYZ] \approx \frac{n-1}{n} \frac{[XY][YZ]}{[Y]} \left((1-\phi) + \frac{\phi N}{n} \frac{[XZ]}{[X][Z]} \right), \quad (4)$$

where N is the total population size. Equation (4) is formed as the number of combinations of $[XY]$ and $[YZ]$ given that the pairs share a common Y site, together with the inclusion of the correlation between X and Z if the three sites form a triangle. This still assumes that the third-moment terms are negligible, but some consideration is given to the structure of the contact network (for a more detailed explanation of equation (4) see Rand (1997)). Henceforth, we use equation (4) to approximate the triple structures in the network.

The extension from consideration of individuals to the modelling of interacting pairs has a strong impact on the dynamics. Although age structure, found to be the main driving force in measles epidemiology (Schenzle 1984; Grenfell *et al.* 1995), has been ignored, it is still possible to reproduce regular biennial epidemics reminiscent of the dynamics in developed countries (Morris 1997). The basic effects of including pair correlations are twofold. An infectious individual can only infect, at most, n other individuals as the disease can only spread to the connected susceptible neighbours; this limits the exponential growth of the epidemic. Each act of infection replaces an S - I pair by an I - I pair, which has a fast dampening effect, it also means that large reservoirs of susceptibles can endure epidemics simply because of the absence of connections.

The pair formulation (equations (2) and (4)) decomposes the familiar transmission β into three components, τ , n and ϕ . The transmissibility τ just measures the rate of an infection spreading between connected individuals, n is the average number of neighbours, and ϕ measures the interconnectedness of the network. This new formulation allows us to express the seasonality more naturally by modifying the network, but keeping the transmissibility τ constant. A

Figure 2. Pictorial representation of a neighbourhood. The main pair X - Y each has nine neighbours; for example, X is connected to three x , five z and Y . X and Y share some of their neighbours (labelled z), which produces triangles in the network (thick lines). For this example $\phi = \frac{5}{8}$.

change to the network will necessitate a change in the number of pairs; in particular, if the average number of neighbours changes to n' , then the number of X - Y pairs changes to $(n'/n)[XY]$.

It is useful to be able to relate the new parameters to the more familiar ones. Summing the pairs containing E in equation (2) and comparing to the SEIR equation (1) we find that

$$\begin{aligned} \beta[S][I] &= \frac{\tau([SSI] + [ESI] + [ISI] + [RSI]) + \tau[SI]}{n} \\ &= \left(\tau[SI] + \tau \sum_X [XSI] \right) / n = \tau[SI]. \end{aligned} \quad (5)$$

As we would expect, this is the rate at which the infection spreads, and it is proportional to the number of susceptible infectious connections. It is reassuring to see that the instantaneous spread of infection is not dependent on any triples (as there are no ϕ terms present), and therefore is independent of the details of the network.

(a) A simplification

Unfortunately, the pair of equations, (2) and (4), are nine dimensional. It would clearly be advantageous if we could use some biological intuition to simplify them by reducing the number of equations. To do this it will be helpful to introduce some additional notation. Let C_{XY} be the correlation between individuals of types X and Y .

$$C_{XY} = C_{YX} = \frac{N[XY]}{n[X][Y]}. \quad (6)$$

In this formulation, $C_{XY} = 0$ corresponds to complete avoidance, $C_{XY} = 1$ to random mixing and values of C_{XY} greater than one to positive correlations.

Equation (5) can now be written in terms of correlations, number of individuals, and neighbourhood size:

$$\beta[S][I] = n\tau[S][I]C_{SI}.$$

We shall make two biologically reasonable assumptions that allow us to reformulate equation (2) just in terms of the number of singles and one correlation, C_{SI} . First, we assume that all correlations C_{XY} other than C_{SE} and C_{SI} (in particular C_{SS}) remain constant in time. Second, we assume that the exposed and infectious periods are short compared to the time that an individual is susceptible. This is the case for measles. Then we can argue that if $\phi = 0$ so that there are no triangles present and, therefore, adjoining pairs are assumed uncorrelated, C_{SE} is well approximated by κC_{SS} , where $\kappa = (n-1)/n$. This is because when a susceptible becomes exposed, this newly exposed individual inherits the neighbourhood of the susceptible. If the infection does not last too long compared to the time that typical neighbours are susceptible, then there is not much time for this correlation to decay. The factor κ enters into this approximation because at least one neighbour of a new infective must also be infected. If $\phi > 0$, some neighbours of a newly formed exposed individual will also be neighbours of the individual that infected it; then, applying this argument and using the ideas of the derivation of equation (4),

$$C_{SE} \approx \frac{n-1}{n} C_{SS} [(1-\phi) + \phi C_{SI}].$$

This unfortunately leads to the inclusion of a new parameter C_{SS} , although its value could be estimated from equation (2).

The nine-dimensional system (equation (2)) can now be replaced by a four-dimensional system (only one more equation than the original SEIR model). Setting $\psi = (SIC_{SI}/N) = ([SI]/n)$ and summing over pairs we obtain equations for the number of individuals in each state.

$$\left. \begin{aligned} \dot{S} &= mN - \tau n\psi - mS, \\ \dot{E} &= \tau n\psi - aE - mE, \\ \dot{I} &= aE - gI - mI, \\ \dot{\psi} &= mI - 2m\psi + aE \frac{S}{N} C_{SE} - g\psi \\ &\quad - \tau[(n-1)\psi + S + \phi(n-1)(C_{II} - 1)\psi] \frac{\psi}{S} \\ &\approx mI - 2m\psi + a \frac{(n-1)}{n} EC_{SS} \\ &\quad \times \left[(1-\phi) \frac{S}{N} + \phi \frac{\psi}{I} \right] \\ &\quad - g\psi - \tau[(n-1)\psi + S] \frac{\psi}{S}. \end{aligned} \right\} \quad (7)$$

This new simplified version requires the estimation of one extra parameter, C_{SS} , compared to equation (2), but has drastically reduced the number of

dimensions. In the approximation $\phi(C_{II} - 1)\psi^2/S$ has been neglected; this is valid as $\psi \ll S$, and as there should be little aggregation of infectious individuals, $C_{II} \approx 1$. Although exposed and infectious individuals must be correlated (as to become exposed an infectious individual must be present), by the time the exposed individual is itself infectious, its source of infection has recovered, so C_{II} should be close to one. The correlation between susceptibles is assumed to be greater than one, as we expect both susceptibles and recovered to be aggregated. In fact, we can see the estimation of C_{SS} to be equivalent in many ways to the addition of age structure. In an age-structured model we would expect pre-school age classes to be predominantly susceptible, and adult classes to be predominantly recovered, which therefore introduces a positive correlation between susceptibles. The dynamics of equations (7) and (2) closely agree for reasonable parameter choices, and the approximations made are accurate.

The only difference between the standard SEIR equations (1) and equation (7) is the addition of an equation which includes the correlation between susceptible and infectious individuals C_{SI} . If $C_{SI} \equiv 1$ then the two equations are identical; but when the correlation is reduced, susceptible and infectious individuals are separated, reducing the size of the epidemic. We shall examine how C_{SI} is affected by various factors:

$$\begin{aligned} \dot{C}_{SI} &= \frac{d}{dt} \left(\frac{\psi N}{SI} \right) = \frac{\dot{\psi} SI - \psi \dot{S} I - \psi \dot{I} S}{S^2 I^2} N \\ &= mIN \left(\frac{SI}{N} - \psi \right) \frac{N}{S^2 I^2} \\ &\quad - \tau \psi I (S - \psi) \frac{N}{S^2 I^2} + a \frac{(n-1)}{n} SE \\ &\quad \times \left[(1-\phi) \frac{SI}{N} C_{SS} + \phi \psi C_{SS} - \frac{n}{n-1} \psi \right] \frac{N}{S^2 I^2}. \end{aligned} \quad (8)$$

For even greater clarity, we shall examine the changes in C_{SI} starting from a virgin state (random mixing, $C_{SI} = 1$):

$$\dot{C}_{SI} = a \frac{E}{I} \left[C_{SS} - 1 - \frac{C_{SS}}{n} \right] - \tau \left(1 - \frac{I}{N} \right). \quad (9)$$

This agrees with our intuitions about the spread of disease. The conversion of exposed individuals to infectious (first term in equation (9)) will usually act to increase the correlation between S and I as the disease is expanding into a new collection of susceptibles which are assumed to be aggregated. The transmission of infection (last term in equation (9)) has the opposite effect, it destroys the S - I pairs through which the disease is transmitted. It should be noticed that initially (when the correlation C_{SI} is one) the structure of the network given by n and ϕ has very little effect. Returning to equation (8) we see that the remaining component, the birth and death of individuals, acts to restore C_{SI} towards one.

Now we understand the role played by ψ in the dynamics of an epidemic, we will look at the long-term behaviour. Typical dynamics of equation (7)

Figure 3. Typical output from the SEIR model in the chaotic regime (grey dashed line) together with regular annual (dotted) and biennial (solid) cycles from the simplified pair model (equation (7)) which do not suffer from the low troughs. The parameters are those given in table 2 and with a population of one million; the annual cycles are produced by $\phi = 0.2$.

Table 1. *Parameter values used in equation (7)*

parameter	value
m^{-1}	50 years
a^{-1}	8 days
g^{-1}	5 days
τ	7.35×10^{-3}
n	$665(1 + 0.28 \cos(2\pi t))$
ϕ	0.08
C_{SS}	2.0

are shown in figure 3, using the parameters given in table 1. If we assume that $\beta = (n\tau/N)$, then the parameters in table 1 correspond to the chaotic regime of the standard SEIR model (Olsen & Schaffer 1990), shown by a grey dashed line on figure 3. The simplified pair model (7), however, can produce biennial cycles (solid line) and annual cycles (when $\phi = 0.2$, dotted line) which do not display such deep troughs as seen in the standard SEIR approach. These models lack the refinements that arise when age structure is included, but demonstrate the strong stabilizing effects that pair correlations confer.

4. FORMULATION OF AN AGE-STRUCTURED MODEL

It is generally accepted that the inclusion of age structure into models for measles is vital to understand and predict the dynamics (Schenzle 1984; Bolker 1993). The resulting realistic age-structured model (RAS) is a considerable improvement on the original SEIR equations, predicting realistic biennial cycles. With this in mind we shall consider an age-structured pair (ASP) model of equation (2) and compare the results to those from the RAS model.

To allow greater flexibility in the ASP model, including the addition of rudimentary family structure, the single set of parameters τ , n and ϕ will be replaced by multiple neighbourhoods, each with a common transmissibility. For example a school child may be a member of two sub-networks, a school sub-network where the number of neighbours and triangular connections are large, and a family sub-network where transmission rates are higher but the number of neighbours are far less. The network within schools may be further subdivided so that heterogeneities in the number of contacts can also be modelled. We shall also need to define the contact structure between, as well as within, age classes. The model is composed of four layers (labelled 1–4), one for each of the age classes: pre-school (0–5 years), primary (6–10 years), secondary (11–15 years) and adults (16+ years). We must now model the behaviour of the 256 pairs (four age classes and four states); although it is not difficult in principle to derive the equations for this system, it is rather lengthy and does not provide new insights. Therefore, we will restrict ourselves to numeric integration of the ASP model.

Each layer obeys the same equations as before, with reproduction and death being restricted to the adult class. However, there are now two transmissibilities used, τ^F measures transmission within a family network and $\tau^R < \tau^F$ measures transmission within the remaining network. The network of contacts can be specified by the number of neighbours of class j that an individual in class i is connected to, labelled n_{ij}^F and n_{ij}^R depending upon whether the connection is to family members or not. To simplify the parameters, we assume that inter-connections are only relevant within an age class, so only ϕ_{ii} are non-zero. This approximation can be justified as the number of connections to outside an age class is generally very small. It should be noted that the parameters n_{22} , n_{33} , ϕ_{22} and ϕ_{33} regulate the size of highly connected groups and therefore the size of primary and secondary schools where most transmission of infection occurs. Typical parameters used are given in table 2.

The pair approach provides an alternative to the RAS model (Schenzle 1984) for the inclusion of age structure. Whereas the RAS model is formulated in terms of the global behaviour, the ASP model is built on an individual-based approach. This both provides advantages and causes complications for parametrization.

(a) Results

A detailed description of all the possible dynamics is not feasible because of the freedom available in the choice of the parameters n and ϕ . However, within the range of reasonable parameters two-year cycles reminiscent of the RAS model (figure 4a), and longer more complex cycles (figure 4b) are possible, all of which correspond well to the England and Wales data set (figure 1).

Table 2. *Parameter values used in the ASP equation*

($\mathbb{P}(n, \phi)$ is the probability of having n neighbours with a connectedness of ϕ . For within each age class all transmissibilities are $\tau^R = 0.2$, between age classes all transmissibilities are $\tau^F = 0.5$.)

from	connected to			
	pre-school	primary	secondary	adult
pre-school	$\mathbb{P}(2, 1) = 0.2$ $\mathbb{P}(4, 0.75) = 0.3$ $\mathbb{P}(6, 0.5) = 0.5$	$\mathbb{P}(0, 0) = 0.5$ $\mathbb{P}(1, 0) = 0.3$ $\mathbb{P}(2, 0) = 0.1$ $\mathbb{P}(4, 0) = 0.1$	$\mathbb{P}(1, 0) = 1.0$	$\mathbb{P}(2, 0) = 1.0$
primary school	$\mathbb{P}(0, 0) = 0.8$ $\mathbb{P}(2, 0) = 0.2$	$\mathbb{P}_{\text{term}}(40, 0.85) = 0.5$ $\mathbb{P}_{\text{term}}(40, 0.5) = 0.5$ $\mathbb{P}_{\text{hols}}(8, 0.6) = 1.0$	$\mathbb{P}(2, 0) = 1.0$	$\mathbb{P}(3, 0) = 1.0$
secondary school	$\mathbb{P}(1, 0) = 1.0$	$\mathbb{P}(2, 0) = 1.0$	$\mathbb{P}_{\text{term}}(45, 0.78) = 1.0$ $\mathbb{P}_{\text{hols}}(15, 0.67) = 1.0$	$\mathbb{P}(5, 0) = 1.0$
adult	$\mathbb{P}(2, 0) = 1.0$	$\mathbb{P}(2, 0) = 1.0$	$\mathbb{P}(2, 0) = 1.0$	$\mathbb{P}(25, 0.4) = 1.0$

Figure 4. Graphs (a) and (b) show the proportion of infectious people in a population, demonstrating some of the widely different dynamics available. Graph (a) has biennial dynamics reminiscent of the RAS model. Graph (b) shows chaotic behaviour but with epidemics still occurring approximately every two years. Despite its chaotic nature there are no very low troughs with this model.

It is found that the majority of the dynamics is within the primary school layer where the fluctua-

tions in the number of cases are the largest. This is in agreement with the age structured data from developed countries. It also appears that when the dynamics are most erratic (figure 4b) it is the slower transmission within the pre-school layer that forms a reservoir of the disease, maintaining the epidemics through the troughs. This would imply that it is the family structure that accounts for the majority of contacts being pre-school and primary school, that enables measles to persist.

(b) *Stochastic effects*

So far the ASP model has only been demonstrated to be comparable to the deterministic RAS model. The existence of parameter regimes displaying complex cycles hardly justifies all the extra work and computation involved in the pair formulation. The deterministic version of the RAS model fits the mean dynamics of England and Wales measles epidemics remarkably well, but a stochastic version reveals discrepancies between the predicted and observed critical community size and levels of fade-out (Bartlett 1957; Black 1966; Grenfell *et al.* 1995; Keeling & Grenfell 1997). We now consider a stochastic ASP model based on the underlying stochastic process. In this the rates for the various possible events are continuously calculated and the events stochastically simulated. If an event occurs at a rate r , then it occurs at a time t into the future with probability re^{-rt} . For greater speed we can use an approximation in which the number of events of a given type within a short time interval is considered to be binomially distributed. In this way one obtains multiple events in a given time period.

Figure 5a gives the average number of fade-outs (defined as three or more consecutive weeks without infection) per year for the parameters of figure 4b, and compares the result to the England and Wales data and to the RAS model. For the biennial parameters (figure 4a) the number of fade-outs is even lower

Figure 5. Results from a stochastic version of the pair model. (a) is the average number of fade-outs per year: the solid line is the best exponential fit to the prevaccination England and Wales data, the circles are the results from a stochastic version of the RAS model, and the crosses are from the pair model. The pair model predicts a lower critical community size, and is a good fit for small populations. (b) shows the power-law behaviour for the size of an epidemic in a small isolated community; population size 25 000, with on average three infectious individuals introduced every four years. There are three distinct sections of power-law scaling, within family, within school, and between school.

with the overall behaviour being consistent. (Introduction of infection in both the ASP and RAS model occurs at a constant Poisson rate of 20 infectious individuals per year per million of population.)

The ASP model predicts a critical community size of around one million, which is far less than is found with the RAS model (or the SEIR model) when the critical community size exceeds 20 million. This can be attributed to the presence of correlations between infectious and susceptible individuals. These correlations have a strong damping effect on the growth rate of infection during the epidemic; also, due to the limited number of contacts, reservoirs of susceptibles are more easily sustained in the population. More striking than the lower critical community size, which could be obtained from large metapopulation models or other modifications to the RAS model (Keeling

& Grenfell 1997), is the agreement between the ASP model and data for small populations. This would indicate that the pair model is capturing the essential local features of epidemics, although it may be failing to reproduce all the larger scale heterogeneities.

Recently, attention has focused on the power-law behaviour of epidemics. Rhodes & Anderson (1996*a, b*) have shown that the size of epidemics in small, isolated island communities and in a simple explicit spatial model obey a power law. In particular, the probability that an epidemic infects more than S individuals ($\mathbb{P}(> S)$) is proportional to S^b , where b is some arbitrary constant. Standard RAS and SEIR models appear to be incapable of generating this form of behaviour; it may be hoped, however, that as the pair system directly models local correlations it too will demonstrate power-law scaling. Using a population of only 25 000 (well below the critical community size) and with a Poisson influx of 0.75 infectious individuals per year, the sizes of epidemics over a 500 year period were recorded.

Figure 5*b* shows $\mathbb{P}(> S)$ against S on a log-log scale. Three distinct regions of power-law behaviour can be identified. For $S < 4$, there is a rapid decline in $\mathbb{P}(> S)$: this is because many incoming infections fail to spread outside the immediate family, and this decline therefore reflects the number of susceptible individuals in the family that contains the infective seed. For $4 < S < 300$ the decline in $\mathbb{P}(> S)$ is extremely shallow. In this region the epidemic has spread into the school environment and the shallow slope indicates that such an infection will normally spread throughout the entire school. In the final region, $S > 300$, we are observing the transmission between schools which is much weaker than the transmission within schools, leading to a sharp decrease in $\mathbb{P}(> S)$.

5. CONCLUSIONS

We have clearly demonstrated the power of correlation models to reflect the local spatial heterogeneities in an epidemic system by considering pairs of connected individuals. From the basic equation (2) we have deduced a simple model (equation (7)) which includes the correlation between I and S , which is the most important in the transmission of an infection. Although this simple model still contains many failings, it provides an alternative to the standard SEIR model, and includes simple heterogeneity.

The more complex age-structured version of equation (2) has been shown to be capable of producing all the dynamics of the RAS model as well as far more complex (possibly chaotic) dynamics. However, it is only when we move to a stochastic framework that the full advantages of the age-structured pair (ASP) model become apparent. As well as producing a lower more realistic critical community size, it mimics the fade-outs for small towns remarkably well.

In contrast to the work of Rhodes & Anderson (1996*a*), our model predicts that for small populations there should be three distinct regions of power-

law behaviour, which have good biological justification. Hence, in epidemiology, and in ecology in general, we should expect to find multiple power-law behaviour, with the relationship in each region relating to the dynamical behaviour at that scale. Correlation models, such as those developed here, are a means to study the dynamics of systems where multiple spatial scales are observed.

This approach obviously has applications to many other communicable diseases. Interesting effects occur in even the most simple systems when correlations are taken into account in this way (Morris 1997). Also, the approach is particularly appropriate for sexually transmitted diseases where the network structure may be far more apparent (Bauch & Rand 1997). As well as considering systems in which the individuals' states are changing, one can also allow the partnership network to change, and to consider the effects of changing such determinants as concurrency and casual partnership frequency. However, much work remains to be done both from a theoretical perspective to determine when correlation models are a good approximation and from an applied approach using details of actual networks to parametrize the models.

We thank Bryan Grenfell and Minus van Baalen for their comments and advice. We also gratefully acknowledge the support of the Wellcome Trust (M.J.K.), NERC (D.A.R.) and EPSRC (D.A.R. and A.J.M.).

REFERENCES

- Anderson, R. M. & May, R. M. 1992 *Infectious diseases of humans*. Oxford University Press.
- Aron, J. L. & Schwartz, I. B. 1984 Seasonality and period-doubling bifurcations in an epidemic model. *J. Theor. Biol.* **110**, 665–679.
- Bartlett, M. S. 1957 Measles periodicity and community size. *J. R. Statist. Soc. A* **120**, 48–70.
- Bartlett, M. S. 1960 The critical community size for measles in the US. *J. R. Statist. Soc. A* **123**, 37–44.
- Bauch, C. & Rand, D. A. 1997 Dynamic networks and correlation equations for STDs. (In preparation.)
- Benenson, A. S. 1990 *Control of communicable diseases in man*, pp. 269–275. American Public Health Association.
- Black, F. L. 1966 Measles endemicity in insular populations: critical community size and its evolutionary implications. *J. Theor. Biol.* **11**, 207–211.
- Bolker, B. M. 1993 Chaos and complexity in measles models: a comparative numerical study *IMA J. Math. Appl. Med. Biol.* **10**, 83–95.
- Bolker, B. M. & Grenfell, B. T. 1993 Chaos and biological complexity in measles dynamics. *Proc. R. Soc. Lond. B* **251**, 75–81.
- Fine, P. E. M. & Clarkson, J. A. 1982 Measles in England and Wales. I. An analysis of factors underlying seasonal patterns. *Int. J. Epidemiol.* **11**, 5–15.
- Grenfell, B. T., Bolker, B. M. & Kleczkowski, A. 1995 Seasonality and extinction in chaotic metapopulations. *Proc. R. Soc. Lond. B* **259**, 97–103.
- Hamer, W. H. 1906 Epidemic diseases in England—the evidence of variability and of persistency of type. *Lancet* **1**, 733–739.
- Hassell, M. P. & May, R. M. 1974 Aggregation in predators and insect parasites and its effects on stability. *J. Anim. Ecol.* **43**, 567–594.
- Johansen, A. 1996 A simple model of recurrent epidemics. *J. Theor. Biol.* **178**, 45–51.
- Keeling, M. J. & Grenfell, B. T. 1997 Disease extinction and community size: modeling the persistence of measles. *Science* **275**, 65–67.
- Matsuda, H., Ogita, N., Sasaki, A. & Satō, K. 1992 Statistical mechanics of population: the lattice Lotka–Volterra model. *Prog. Theor. Phys.* **88**, 1035–1049.
- Morris, A. 1997 Representing spatial interactions in simple ecological models. Ph.D. thesis, University of Warwick, UK.
- Olsen, L. F. 1987 Low dimensional strange attractors in epidemics of childhood diseases in Copenhagen, Denmark. In *Chaos in biological systems*, pp. 249–254. London: Plenum.
- Olsen, L. F. & Schaffer, W. M. 1990 Chaos versus noisy periodicity: alternative hypotheses for childhood epidemics. *Science* **249**, 499–504.
- Rand, D. A. 1997 Correlation equations for spatial ecologies. *Theoretical ecology* (ed. J. McGlade), vol. 2. Oxford: Blackwell.
- Rand, D. A. & Wilson, H. B. 1991 Chaotic stochasticity—a ubiquitous source of unpredictability in epidemics. *Proc. R. Soc. Lond. B* **246**, 179–184.
- Rhodes, C. J. & Anderson, R. M. 1996a Power laws governing epidemics in isolated populations. *Nature* **381**, 600–602.
- Rhodes, C. J. & Anderson, R. M. 1996b A scaling analysis of measles epidemics in a small population. *Phil. Trans. R. Soc. Lond. B* **351**, 1679–1688.
- Satō, K., Matsuda, H. & Sasaki, A. 1994 Pathogen invasion and host extinction in lattice structured populations. *J. Math. Biol.* **32**, 251–268.
- Schenzle, D. 1984 An age-structured model of pre- and post-vaccination measles transmission. *IMA J. Math. App. Med. Biol.* **1**, 169–191.
- Schwartz, I. B. 1985 Multiple recurrent outbreaks and predictability in seasonally forced nonlinear epidemic models. *J. Math. Biol.* **18**, 233–253.

Received 17 April 1997; accepted 15 May 1997