CHAPTER 10

Dynamics of infectious disease

Bryan Grenfell and Matthew Keeling

10.1 Introduction

Host–pathogen associations continue to generate some of the most important applied problems in population biology. In addition, as foreshadowed in Chapter 5 of this volume, these systems give important insights into the dynamics of host–natural enemy interactions in general. The special place of pathogens in the study of host–natural enemy dynamics arises partly from excellent long-term disease-incidence data, reflecting the public health importance of many infections. However, we argue that host–pathogen dynamics are also distinctive because the intimate association between individual hosts and their pathogens is often reflected with particular clarity in the associated population dynamics. Throughout this chapter we focus in parallel on the population dynamics of host–pathogen interactions and the insights that host–pathogen dynamics can provide for population biology in general.

Population-dynamic studies of infectious disease have a long history, which predates the modern foundations of ecology (Bernoulli, 1760). During the twentieth century, the preoccupation of population ecologists with the balance between extrinsic and intrinsic influences on population fluctuations and the role of nonlinearity and heterogeneity (Bjørnstad and Grenfell, 2001) find strong parallels in epidemiological studies of human diseases (Bartlett, 1956; Anderson and May, 1991). In terms of the ecological effects of parasitism, the traditional view held that 'well-adapted' parasites would not have a consistent impact on the ecology of their hosts (Grenfell and Dobson, 1995). The 1970s saw a new departure, when Anderson and May pointed out the potential of infectious agents to exert nonlinear—regulatory or destabilizing—influences on the population dynamics of their hosts (Anderson and May, 1978, 1979; May and Anderson, 1978, 1979). There has since been an explosion of work on the population biology of human, animal, and plant pathogens. This work spans a huge range: from highly applied to basic theoretical work; from within-host to the metapopulation scale; from short-term population dynamics to long-term evolutionary processes. In this chapter we first outline the simple theory of epidemiological models; we then refine this picture to illustrate the potential impact of pathogens on the population dynamics of their hosts, as well as aspects of host–pathogen interactions which provide important insights into more general ecological dynamics. A particular focus of this review is the calibration of simple and complex epidemiological models against the rich databases available for many host–parasite systems.

Before the 1970s, most theoretical studies in epidemiology focused on the dynamics of epidemics in constant host populations. Although this approximation is reasonable for acute infections in human populations, it will not do if we wish to consider the impact of infectious disease on the dynamics of natural populations. This issue was addressed by Anderson and May (Anderson and May, 1979; May and Anderson, 1979), who proposed a unified framework for capturing the dynamics of parasites and their impact on the ecology of their hosts. They divided parasites into two broad groups: microparasites (mainly viruses, bacteria, and other micro-organisms) and macroparasites (mainly helminths and arthropods). Despite their names, the main differentiation between the two groups is a functional one

derived from their life history. For microparasites, which reproduce rapidly within the host, the resulting theoretical framework is based upon keeping track of the number of *hosts* of different types (e.g. susceptible to infection, infected, or recovered from infection). In contrast for macroparasites, which generally reproduce more slowly via multiple transmission stages, it is more important to model the number of *parasites*.

The study of the population dynamics of infectious diseases has seen several major advances over the past decades (Anderson and May, 1991; Hudson *et al.*, 2001). In brief, we can partition these developments into six basic categories, each of which brings epidemiological models closer to reality.

1 Host heterogeneities. The most basic disease models assume that all individuals are identical, with equal risks of infection and subsequent transmission. However, motivated by observations of sexually transmitted infections, it was realized that the heterogeneity in the number (and type) of contacts could play a crucial role in the transmission dynamics—with a highly connected core group responsible for maintaining the infection (Hethcote *et al.*, 1982; Anderson and May, 1991; Morris, 1993; Lloyd-Smith *et al.*, 2005).

2 Within-host dynamics. The basic compartmental models for infectious diseases treat all infected hosts equally—leading to a constant transmission rate throughout an exponentially distributed infectious period. In reality, the transmission rate varies during the infectious period, which itself is better approximated as a normal or gamma distribution (Keeling and Grenfell, 1997). Such details of pathogen biology can have a profound effect on epidemic dynamics (Grossman, 1980) and the impact of control mechanisms (Fraser *et al.*, 2004; Wearing *et al.*, 2005). This work has strong parallels with metapopulation theory in ecology, with simple compartmental models being equivalent to Levins metapopulations, whereas more complex models describe the 'within-patch' dynamics. Capturing the within-host behaviour can itself be seen as a question of population dynamics (Nowak and May, 2000); there is currently substantial interest in modelling the interaction of the pathogen and the host's immune system, how this generates the observed host-level behaviour, and how within-host dynamics translate to the population level (Grenfell *et al.*, 2004).

3 Parasite genetic heterogeneities and evolutionary dynamics. Microparasite evolutionary dynamics are some of the most spectacular manifestations of host–pathogen heterogeneity, as well as particularly clear examples of rapid evolution in action. The evolution of parasite drug resistance and immune escape also represent some of the major problems in the fight against disease (Frank, 2002). We return to these issues below. Another major challenge is to understand the complex interplay between host and parasite genetic heterogeneity, and interpret how this will be reflected in the epidemiological and evolutionary dynamics.

4 Temporal forcing. The transmission of many pathogens is affected by external forcing at seasonal and other temporal scales. Seasonal effects have been particularly well characterized in the transmission of childhood infections, where school terms and holidays impact upon the level of transmission; this forcing in turn interacts with the host–pathogen epidemic dynamics producing the potential for rich and complex dynamical behaviour (Schaffer and Kot, 1985a, 1985b; Olsen *et al.*, 1988; Rand and Wilson, 1991). School terms are not the only source of forcing; more generally, climatic fluctuations influence transmission, leading to long-period oscillations (for example, the influence of the El Niño southern oscillation on cholera dynamics; Koelle *et al.*, 2005).

5 Spatial structure. A fundamental feature of most infectious disease transmission is that it only occurs between individuals in 'close' contact. Therefore diseases tend to spread locally in space, with proximity acting as a major risk factor (Bartlett, 1960; May and Anderson, 1984; Cliff and Haggett, 1988; Murray, 1989; Cliff *et al.*, 1993; Grenfell *et al.*, 2001; Smith *et al.*, 2002). Again this work has clear parallels with ecological dynamics where the locality of dispersal can have profound effects (see Chapters 4 and 5 in this volume). More recently attention has focused on the network structure of contacts, explicitly modelling the local interactions that can lead to disease transmission (Keeling and Eames, 2005).

6 Stochasticity. The random transmission of infection and the discrete nature of the host population has a fundamental, and often very relevant, influence on disease dynamics. Stochastic models provide us with a method of capturing these two essential elements. As such, demographic stochasticity is most important when dealing with low numbers of cases—when it can give rise to local extinctions of the disease (Bartlett, 1960; Keeling and Grenfell, 1997). As described below, stochasticity can have a major impact on the dynamics of both micro- and macroparasites.

All these refinements allow us to capture more of the known dynamical behaviour of an infectious disease. However, a recurring theme in the chapter is that simple models are often remarkably successful at generating robust dynamical insights, which can inform more detailed models.

We structure our tour of host–parasite population dynamics by considering progressively more realistic assumptions about host ecology and the biological details of the host–parasite interaction. We begin by outlining a basic model for the dynamics of a microparasite invading a closed and constant host population. Though simple, this scenario allows us to derive the crucial thresholds for parasite invasion and spread, which apply universally across more ecologically realistic situations. We then explore how childhood infections epitomize the nonlinear spatio-temporal dynamics of ecological population cycles. Turning to the ecological impact of parasitism, we revisit classical models for the impact of parasites on host-population regulation and the special dynamical role of heterogeneities in the host–parasite relationship. A concluding case study on foot and mouth disease illustrates a range of other features of host–pathogen dynamics, particularly the synthesis of simple and complex models of population dynamics.

10.2 Microparasite models: the simple epidemic

The essential functional feature of microparasites is that they reproduce directly within their hosts (Anderson and May, 1991). Thus, for many diseases, we can capture the essential dynamics via a *compartmental model*, where hosts are divided between different infection categories. The most studied form is the family of susceptible → infected → recovered (SIR) models for strongly immunizing infections. Individuals are recruited into a previously uninfected susceptible class, S (often via births). They then become infected (I) by contact with infected individuals, before finally moving into an immune-recovered class, R (Bartlett, 1956; May, 1980; Mollison, 1995). The simplest realization of the SIR model describes the dynamics of the *simple epidemic*—an outbreak of a non-fatal infection directly transmitted between hosts in a closed population, without host demographic changes (May, 1980):

$$\begin{aligned} \frac{dS}{dt} &= -\beta IS \\ \frac{dI}{dt} &= \beta IS - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \qquad (10.1)$$

The model assumes a well-mixed population, with homogeneous random mixing between individuals. As with the simplest predator–prey models (see Chapter 3 in this volume), this leads to a bilinear ($I \times S$) interaction term for the infection process, controlled by the infection parameter, β. This form assumes that transmission depends upon the density of susceptible and infectious individuals (as discussed further below; Begon et al., 2002). For simplicity (eg Bjornstad and Grenfell (2002)) we allow β to subsume the rate at which susceptible and infectious individuals make sufficiently close contact to allow a chance of transmission, and the probability of transmission of the infectious agent given such a contact. After infection, individuals move into the recovered class, R, at the recovery rate, γ.

A key epidemiological parameter leaps out of this model: consider an epidemic sparked by a single infected individual introduced into a total susceptible population of density N. At the start of the epidemic (when $S \cong N$) the level of infection, I, can only increase ($dI/dt > 0$) if $\beta N > \gamma$; this leads naturally to a definition of the *basic reproduction ratio* or *basic reproductive number of infection*:

$$R_0 = \beta N/\gamma \qquad (10.2)$$

Since $1/\gamma$ is the average duration of infection, R_0 can be interpreted as the total number of secondary cases produced by one infected individual when introduced into a population of N susceptibles. As the epidemic proceeds, and more individuals recover and are therefore immune, the effective reproduction ratio, R, declines with the proportion of susceptibles, $s = S/N$; that is, $R = R_0 s = R_0 S/N$. The basic and effective reproduction ratios provide a powerful framework for exploring the dynamics and control of epidemics. In particular, an infectious agent can only invade a susceptible population if $R_0 > 1$; if this criterion is satisfied the subsequent epidemic will only continue to increase while $R > 1$; when the proportion of susceptibles is reduced by the epidemic below $1/R_0$, the number of cases must start to decline. An endemic infection, steadily maintained at a constant level, will have an effective reproduction ratio $R = 1$; but $R = R_0 s^*$, where s^* is the equilibrium fraction who are susceptible, whence R_0 can be estimated as $R_0 = 1/s^*$.

10.2.1 Control by vaccination

In essence, the simple epidemic extinguishes itself by reducing its supply of susceptibles below $1/R_0$. The same threshold also applies to control of immunizing pathogens by vaccinating with an inactivated or attenuated pathogen (Anderson and May, 1991); reducing the proportion of susceptibles below $1/R_0$ will keep R below unity and prevent invasion of the infection. Note that the resulting critical proportion of susceptibles that require vaccination to prevent an epidemic ($p_c = 1 - 1/R_0$) is less than unity; this reflects the concept of *herd immunity* in which not all susceptible individuals in a population need to be protected to prevent a large-scale epidemic—logistical constraints then determine whether this level of vaccination can be achieved.

10.3 Host vital dynamics and long-term behaviour of the SIR model

The simple SIR epidemic described above seals its own fate by depleting its susceptible fuel. However, if we allow for the added realism (in most populations) of a steady replenishment of susceptible individuals via births (Anderson and May, 1991), further, *recurrent* epidemics become possible. This brings us to the SIR model with vital dynamics:

$$\begin{aligned} \frac{dS}{dt} &= \mu(N - S) - \beta IS \\ \frac{dI}{dt} &= \beta IS - (\mu + \gamma)I \\ \frac{dR}{dt} &= \gamma I - \mu R \end{aligned} \qquad (10.3)$$

Again, the total host population size is constant and infection is not assumed to affect host survival or reproductive rates. However, this model replaces the previous assumption of a 'closed' host population, without vital dynamics, by assuming 'background' birth and death rates of hosts, controlled by a per-capita death rate, μ. Although eqn 10.3 is a general formulation, we focus in particular on 'acute' infections, such as measles, where the duration of infection is relatively brief (i.e. $\gamma \gg \mu$). A numerical simulation with measles parameters (Figure 10.1a) immediately illustrates the impact of susceptible replenishment by births—a series of recurrent epidemic oscillations (with an approximately 2-year period for measles parameters), slowly damping to a stable equilibrium abundance of susceptible and infected individuals (May, 1980). Essentially, the nonlinear feedback between infected and susceptible densities, along with the upper limit on susceptible abundance, N, stabilizes their interaction. As shown by the S and I nullclines for the model (Figure 10.1b), these dynamics are closely analogous to the dynamics of Lotka–Volterra predation models with density-dependent limitations on prey abundance (Rosenzweig and MacArthur, 1963).

The period of these damped model oscillations close to equilibrium captures the tendency for measles epidemics to be strongly biennial in the era before vaccination in developed countries (Bartlett, 1957; Anderson *et al.*, 1984; see Figure 10.3a, below). However this comparison of model and data also illustrates a significant question: what additional biological feature do we need to translate the weakly damped epidemics of the

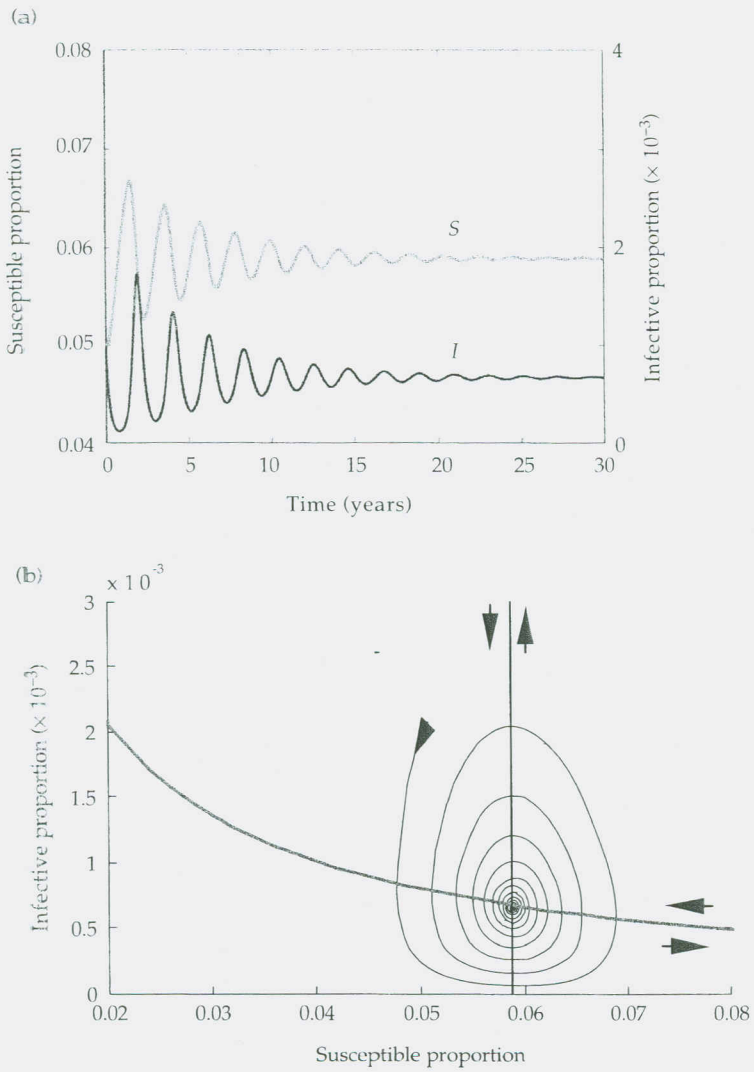

Figure 10.1 Deterministic behaviour of the SIR model including vital dynamics; we use epidemiological parameters for measles (Earn *et al.*, 2000). (a) A numerical simulation showing the damping of susceptible (S, black line) and infected (I, grey line) proportions towards their deterministic equilibrium values. (b) Corresponding nullclines (zero growth contours) for I and S at $S_0 = (\mu + \gamma)/\beta$ and $I_0 = (N - S_0)/\beta S_0$; the arrows show direction vectors for S (black line) and I (grey line) and the spiral corresponds to the simulation in (a).

model into the dramatic sustained oscillations of the observed series?

10.3.1 Seasonal forcing and sustained epidemic oscillations

The above question parallels a long-standing population-dynamic debate about the cause of sustained host–natural enemy cycles of small mammals, game birds, and forest insects (Bjørnstad and Grenfell, 2001). The key issue in ecology is generally to explain the rarity of such cycles in practice compared with basic theory; here, the issue is reversed: how do we explain the violent cycles of measles and other childhood diseases, given the long-term stability of very plausible simple models? As with ecology, much of the debate has revolved around the balance

between intrinsic nonlinear dynamics, time delays, and external forcing (Bartlett, 1956, 1957, 1960; Dietz, 1976; May, 1980; Fine and Clarkson, 1982; Schenzle, 1984; Bolker and Grenfell, 1993).

For measles, Bartlett's key insight is that repeated external perturbations, either from stochastic fluctuations or seasonal variations in transmission, can throw the damped oscillations of the SIR model into sustained epidemic cycles (Bartlett, 1956, 1957, 1960). In practice, seasonal variations in infection rates appear to be the major force sustaining epidemic cycles of measles which are relatively invariant to population size and therefore stochastic effects (Fine and Clarkson, 1982; Schenzle, 1984; Bolker and Grenfell, 1993).

We illustrate this role of seasonality in Figure 10.2. Figure 10.2a shows the dynamics of our simple SIR model with vital dynamics, with the addition of an annual sinusoidal swing of 20% in the per-infective infection rate, β (inset). This repeated perturbation to the system pushes the dynamics into sustained cycles; the deterministic trajectory loops around the unforced equilibrium in *SI* phase space (Figure 10.2b). Furthermore, the annual period of the forcing resonates with the unforced biennial (but damped) tendency to generate focused, violent epidemics reminiscent of the observed pre-vaccination case-report data (compare Figures 10.2a and 10.3a). More broadly, the synthesis of models with data on childhood epidemics illustrates key principles of ecological dynamics in two parallel, and ultimately interlocking, directions.

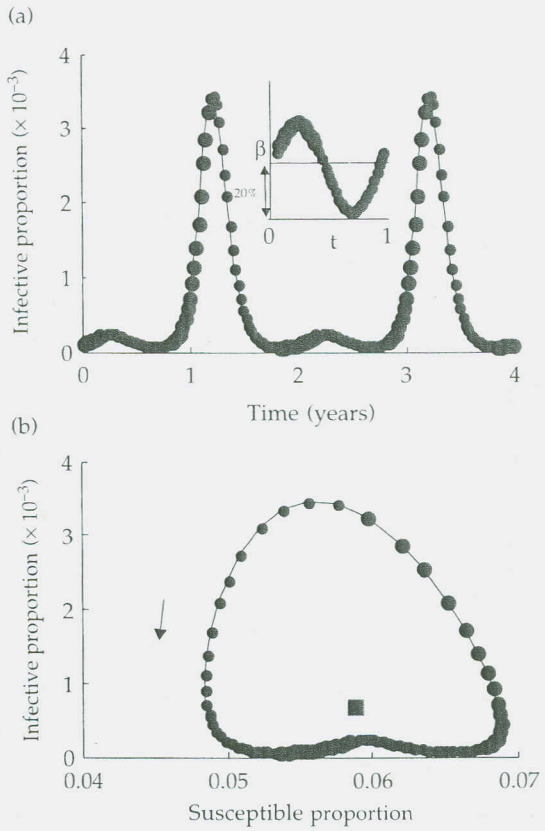

Figure 10.2 Impact of seasonal forcing of the infection rate, β on the dynamics of the simple SIR model, with vital dynamics (Figure 10.1). We use a simple sinusoidal annual forcing: $\beta(t) = \beta_0(1 + \beta_1 \cos(2\pi t))$ (t is time in years), with seasonal forcing amplitude, $\beta_1 = 0.2$ (a 20% seasonal swing). (a) Time series, with the seasonal forcing pattern shown in the inset; symbol area is proportional to forcing amplitude. (b) *SI* phase plane plot corresponding to (a); the arrow shows the direction of trajectories and the square is the unforced equilibrium, (S^*, I^*).

10.3.1.1 Realistic epidemiological models

A major body of work has developed increasingly accurate models for assessing the performance of vaccination strategies and exploring spatio-temporal dynamics (Bartlett, 1956; May and Anderson, 1984; Schenzle, 1984; Anderson and May, 1991; Cliff *et al.*, 1993; Babad *et al.*, 1995; Grenfell *et al.*, 2001). These range from complex age-structured models (Schenzle, 1984; Anderson and May, 1991; Bolker and Grenfell, 1993; Babad *et al.*, 1995)—key for designing age-targeted vaccination strategies—to much simpler aggregate formulations (Earn *et al.*, 2000; Bjørnstad *et al.*, 2002). In fact, simple models (combining epidemic dynamics with realistic seasonal forcing functions) can capture long-term measles dynamics in large cities surprisingly well (Figure 10.3a), as long as we allow for secular changes in birth rates (which effectively tune the propensity for oscillations of different periods; Bjørnstad *et al.*, 2002; Grenfell *et al.*, 2002). As shown in Figure 10.3a, dynamics in these large centres are relatively insensitive to demographic stochastic perturbations. In smaller populations, below a critical community size of around 300 000 individuals, infection generally disappears in the troughs between epidemics due

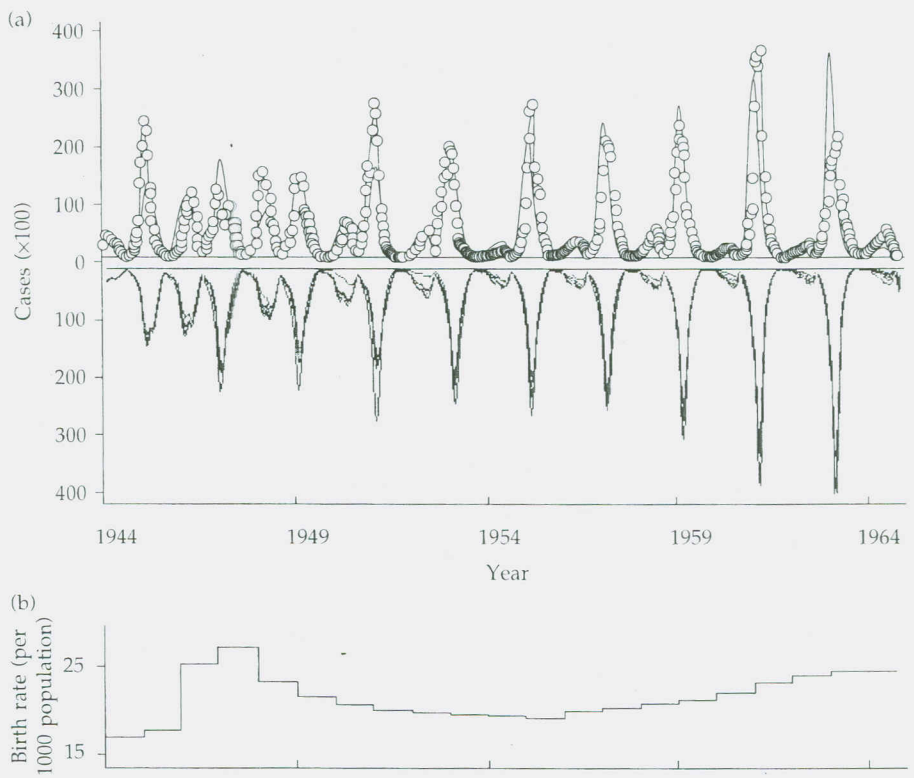

Figure 10.3 (a) Observed biweekly measles notifications for London (corrected for under-notification) in the pre-vaccination era (circles). The line shows a corresponding deterministic simulation of a forced SIR model set in a time-series framework (the TSIR model; see Bjørnstad *et al.* (2002) and Grenfell *et al.* (2002) for details and fitting procedures). The lower reflected graph in (a) shows 10 replicate stochastic simulations of the same model. This underlines that demographic stochasticity has remarkably little effect on the endemic limit-cycle behaviour of measles in large communities. (b) Corresponding annual birth rates (per thousand population) for London; from the 1950s onwards we see sustained biennial cycles; by contrast, the baby boom during the 1940s increased the recruitment of susceptible individuals, driving annual measles epidemics.

to demographic stochasticity (Bartlett, 1960). Another body of literature has used stochastic models to explore these dynamics and the spatio-temporal patterns which emerge from them (Lloyd and May, 1996; Grenfell *et al.*, 2001).

10.3.1.2 Nonlinear dynamics and chaos in epidemics

The search during the 1980s and 1990s for the imprint of chaotic dynamics in real ecological systems (May, 1976a; see also Chapter 3 in this volume) quickly revealed that few ecological systems had the necessary length and detail of time series to discern the sensitivity to initial conditions that is their characteristic fingerprint (Sugihara and May, 1990). In the mid 1980s W.M. Schaffer realized that the relatively long and densely sampled time series of measles, and other strongly fluctuating childhood infections, might provide the key to this enigma (Schaffer and Kot, 1985a, 1985b; Olsen *et al.*, 1988). Schaffer and Kot realized that acute, self-immunizing, seasonally driven infections like measles are also especially suitable candidates for exotic dynamics—their natural oscillatory behaviour has the propensity to become chaotic at relatively high amplitudes of seasonal forcing. The resulting syntheses of models and data provided a great stimulus to our understanding of nonlinear dynamics in ecology (see Chapters 3 and 5 in this volume).

The effects of high-amplitude seasonal forcing on the deterministic dynamics of measles models are illustrated in Figure 10.4. As forcing amplitude

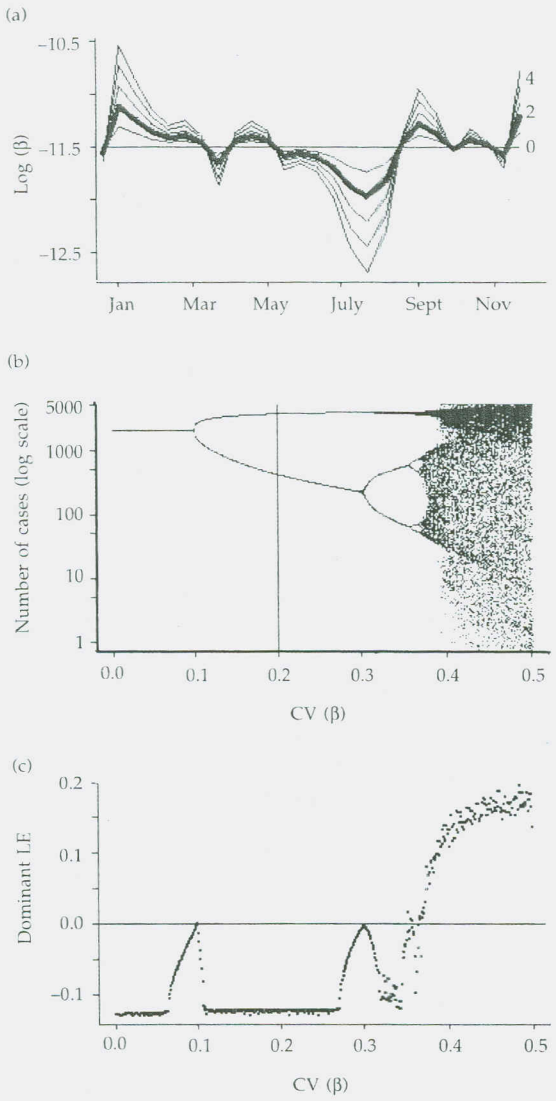

Figure 10.4 Nonlinear behaviour of the SIR model as a function of the amplitude of seasonal forcing. We illustrate the effects of increasing seasonality using the TSIR model (full details of model and figure are given by Grenfell *et al.*, 2002). (a) Observed seasonal variation in infection rate, *I*, for pre-vaccination London (bold line); other lines show increasing amplitudes of forcing used in (b) and (c). (b) Bifurcation diagram for the TSIR model as a function of increasing forcing amplitude (measured as the coefficient of variation (CV) of the patterns in a). For each forcing amplitude, we sampled annually at the start of each year to generate the points. The diagram clearly shows forcing thresholds at approximate coefficients of variation of 0.1 and 0.4, above which the dynamics move to biennial and chaotic behaviours respectively. (c) Corresponding Lyapunov exponents (LE), showing a transition to positive exponents in the chaotic region (Olsen *et al.*, 1988; Tong, 1990; Grenfell *et al.*, 2002).

increases, measles first enters a biennial regime, which dominates over a wide region of forcing/ levels (Figure 10.4b). At high levels of forcing, this behaviour gives way to a region of high-amplitude chaos, characterized by irregular dynamics with deep inter-epidemic troughs. However, in terms of the comparison between models and data, there is still no compelling

empirical evidence for chaos in epidemics, for two reasons. First, from the data perspective, non-stationarity in driving parameters such as birth rate (Figure 10.3) and the underlying stochasticity of transmission effectively reduce the ability to detect the signature of sensitivity to initial conditions (Bjørnstad and Grenfell, 2001; Bjørnstad *et al.*, 2002). Second, the strength of seasonal forcing required for chaotic dynamics in models is stronger than that estimated from the case-reporting data. In addition, the model dynamics within the chaotic parameter region tend to generate deeper troughs with more-localized extinctions of infection—and therefore a much higher critical community size—than are seen in practice (Bjørnstad *et al.*, 2002). Nevertheless, the combination of over-compensatory dynamics, forcing, and noise seen in SIR-type infections make measles, and other acute, strongly forced infections, strong candidates for complex dynamics (Rand and Wilson, 1991; Earn *et al.*, 2000; Keeling *et al.*, 2001a).

10.4 Impact of parasites on host dynamics and population regulation

The last two decades have seen an explosion of empirical evidence on the impact of emerging and established pathogens on the dynamics and evolution of their host populations (Grenfell and Dobson, 1995; Hudson *et al.*, 2001). A significant impetus for this work was the basic theory which elucidated the potential nonlinear impact of micro- and macroparasites on population regulation of their hosts (Anderson and May, 1978; May and Anderson, 1978). To illustrate the general point, we turn to macroparasitic worms, where theory and empirical ecology come together in a particularly elegant way. The key functional properties of macroparasites are that parasites do not reproduce directly in their hosts and generally produce deleterious effects in proportion to their abundance; thus we have to track the intensity of infection in individual hosts. To test the potential for population regulation, Anderson and May (see also Crofton, 1971; May, 1977b) formulated a basic model for a directly (i.e. non-vectored) macroparasite in an exponentially growing host population:

$$\frac{dH}{dt} = (a-b)H - (\delta+\alpha)\overline{m}H$$
$$= (a-b)H - (\delta+\alpha)P \qquad (10.4)$$

$$\frac{dP}{dt} = \frac{\lambda PH}{H_0+H} - b\overline{m}H - vP - \alpha HE(m^2)$$
$$= \frac{\lambda PH}{H_0+H} - (b+v)P - \alpha HE(m^2) \qquad (10.5)$$

Here, a host population (H), with per-capita birth and death rates a and b, respectively, will grow exponentially ($a > b$) in the absence of parasitism. Adult parasites (total population P) produce transmission stages into the environmental at a per-capita rate, λ; the nonlinear success of transmission is captured by the parameter H_0 in eqn 10.5. Additionally, parasites have a background death rate of v. The impact of parasites on the host is assumed to be proportional to the average parasite burden ($\overline{m} = P/H$): those hosts with more parasites suffer a greater reduction in survival and reproduction governed by per-capita rates α and δ respectively.

10.4.1 Heterogeneity and regulation

The final term in both the host and parasite equation (eqn 10.5) allows for two biological processes (Anderson and May, 1978; May and Anderson, 1979): (1) the net host mortality due to parasitism is proportional to average parasite burden ($\overline{m}$), and (2) if parasites kill their host then they themselves die. This combination of effects means that host death can have a disproportionately big effect on parasite mortality, as captured by the expectation $E(m^2)$: the second moment (mean^2 + variance) of the distribution of parasites per host. Therefore, those hosts with a greater-than-average burden are more likely to die, which in turn leads to the death of a greater-than-average number of parasites. In the overwhelming majority of situations (Anderson and May, 1978; Roberts *et al.*, 1995) the parasite

distribution is highly aggregated (variance > mean), which greatly magnifies this effect. To an excellent empirical approximation, parasite numbers per host can be assumed to be negatively binomially distributed with dispersion parameter k and variance $= \bar{m} + \bar{m}^2/k$, where $\bar{m}$ is the mean parasite burden per host ($\bar{m} = P/H$). The negative binomial has second moment $E(m^2) = \bar{m} + \bar{m}^2(k+1)/k$, thus eqn 10.5 becomes (May and Anderson, 1978):

$$\frac{dP}{dt} = \frac{\lambda PH}{H_0 + H} - (b + v + \alpha)P - \alpha\frac{(k+1)P^2}{kH} \qquad (10.6)$$

Ignoring for the moment the reduction in host reproduction due to parasites ($\delta = 0$), Anderson and May show that parasites can regulate the host population to a stable equilibrium as long as the parasite frequency distribution is aggregated (i.e. $k < \infty$; where $k = \infty$ yields a Poisson distribution). Again, this result has general ecological resonance—heterogeneity in encounter rates can contribute significantly to the stability of host–natural enemy interactions (see Chapter 5 in this volume). We return to the special impact of heterogeneities on host–pathogen dynamics in the case study on foot and mouth disease (section 10.5).

10.4.2 Macroparasites and host population cycles

A variety of refinements have been made to the basic host–macroparasite model (Roberts *et al.*, 1995). The most interesting synthesis of theory and field epidemiology arises from the potentially strongly destabilizing effects of reductions in host reproduction rates due to parasitism (i.e. $\delta > 0$). We can immediately discern the qualitative effects of reproductive limitations by parasitism if we re-express the parasite equation (eqn 10.6) in terms of parasites per host; $\bar{m} = P/H$:

$$\frac{d\bar{m}}{dt} = \frac{\lambda \bar{m} H}{H_0 + H} - (a + v + \alpha)\bar{m} - (\alpha/k - \delta)\bar{m}^2 \qquad (10.7)$$

Note that, as we are now dealing with the average number of parasites per host, the natural loss of hosts (at rate b) is irrelevant, whereas the birth of uninfected hosts (at rate a) dilutes the average burden.

Focusing on the quadratic term in $\bar{m}$, parasite-induced host mortality (α) exerts a regulatory negative feedback in an aggregated parasite population ($k < \infty$). However, host reproductive limitations by the parasite (δ) reduce this regulatory effect; indeed, for $\delta > \alpha/k$, the quadratic term becomes positive and the system is thrown into limit cycles (May and Anderson, 1978). Heuristically, reproductive limitations by the parasite reduce host-population growth rate, but have much less strong short-term effects on parasite abundance per host—hence the reproductive effects continue until the host population reaches a lower threshold where the parasite population crashes and the hosts recover, etc. These very distinctive feedbacks again illustrate the special ecological dynamics that can emerge from the intimate relationship between host and parasite.

Probably the best empirically characterized impact of parasitism on long-term host dynamics arises in the ecology of a macroparasite: the nematode *Trichostrongylus tenuis* infecting red grouse populations in northern England and Scotland. Game bag records for the grouse show regular 5–7-year cycles since the nineteenth century. In an echo of other debates about the role of natural enemies in ecological cycles, *T. tenuis*, which has strong effects on host recruitment at high host densities, has long vied with intrinsic behavioural explanations as a candidate for causing the cycles (Hudson *et al.*, 1998, 2001). Because birds can be treated with antihelmintic drugs, parasites are more easily 'excluded' from the system than predators usually are in other cycling host populations. This led Hudson *et al.* (1998) to carry out a decade-long field experiment demonstrating that cycles were prevented on treated grouse moors compared with untreated control moors where the cycles continued. The debate about grouse cycles continues (Mougeot *et al.*, 2005). However, the work of Hudson *et al.* remains one of the best experimental tests of simple theory in population dynamics.

10.5 Foot and mouth disease in the UK

The 2001 foot-and-mouth epidemic had a huge and lasting impact on the UK livestock industry. However, the epidemic also precipitated a substantial change in the way that models of infectious disease are formulated and utilized. We now briefly review the epidemic situation in 2001 and the biological characteristics of foot and mouth disease before showing how these influenced the choice of modelling approach. Apart from its applied importance, the UK foot-and-mouth epidemic illustrates a number of general points about the spatio-temporal dynamics of infectious diseases, as well as the level of model detail required to address different questions.

Foot and mouth disease is one of the most rapidly transmitted of all livestock infections, and can quickly spread both within and between farms. Foot and mouth disease can infect most livestock species, although during the 2001 epidemic in the UK it was primarily cattle and sheep farms that were most frequently affected. In total, 2026 farms reported infection with foot and mouth disease (leading to the culling of their animals), while approximately four times that number of farms had their animals culled in an attempt to stop the spread of infection. Despite this seemingly intense culling, the epidemic lasted from February until September, with a long protracted tail once the main bulk of the epidemic had died away. Figure 10.5 shows the pattern of cases and culls in space and time.

We now outline one of the models used during the 2001 epidemic (Keeling *et al.*, 2001b), focusing on how the epidemiological characteristics dictated the model structure and how models are necessary to interpret the complex nonlinear trade-offs that occur with control. The data from the UK foot-and-mouth epidemic have presented modellers with an astoundingly detailed data-set, with a large amount of information on both susceptible and infected farms, as well as a range of complex

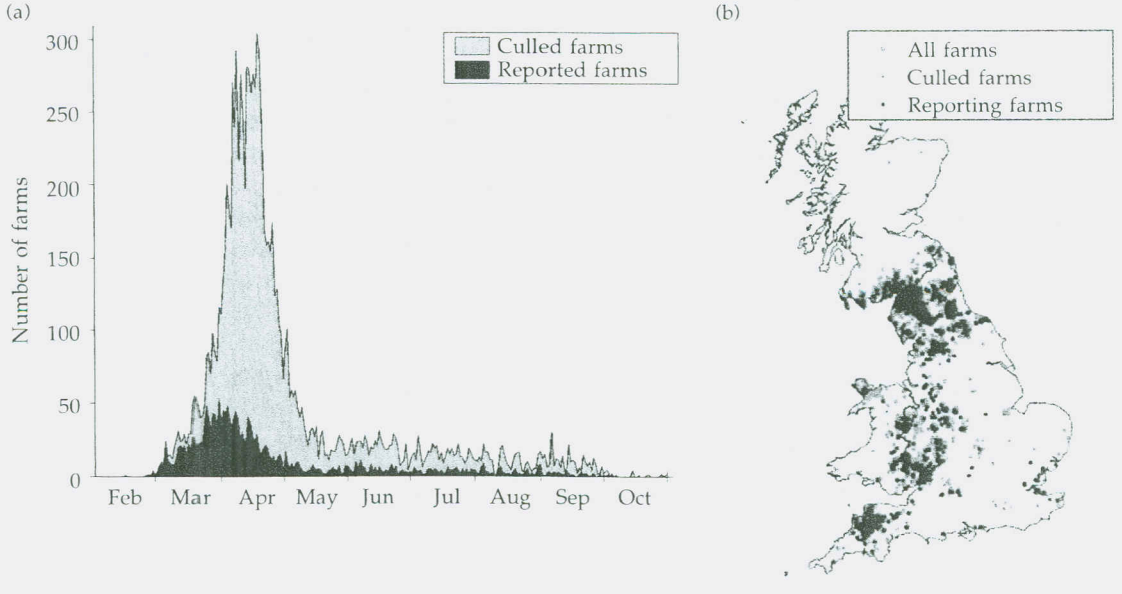

Figure 10.5 The recorded spatio-temporal distribution of cases and culls during the UK foot-and-mouth epidemic in 2001. (a) Time series of daily reported cases (black) and farms whose animals were culled as a preventative measure (grey). It is clear that the majority of farms lost was due to preventative culling, although an uncontrolled epidemic would have been far worse. (b) The spatial location of all farms (pale grey), culled farms (dark-grey circles), and farms reporting infection (black squares). The concentration of infection in Cumbria, Dumfries and Galloway, Devon, and the Welsh borders is clear (Data from DEFRA; www.defra.gov.uk/footandmouth/cases/fmdcases/map.htm).

questions. Three models were utilized during the 2001 foot-and-mouth epidemic (Keeling, 2005); they ranged from elaborate simulations (Morris *et al.*, 2001) to a system of (complex) differential equations (Ferguson *et al.*, 2001a, 2001b). These models generally agreed in the type of additions to the simple SIR models (eqn 10.1) that were required to capture the observed dynamics of foot and mouth disease.

Host heterogeneities. Variability at the host level comes from two different sources, the first is the intrinsic differences between different livestock species. In particular, cattle are far more susceptible than sheep, whereas pigs featured rarely in this epidemic; in addition, cattle are also responsible for a greater amount of transmission than sheep. From this perspective cattle can be seen as a core group, being both more at risk of infection and subsequently a greater potential for generating further cases (Anderson and May, 1991). This heterogeneity is amplified by the clustering of animals (often of the same species) within farms—essentially corresponding to assortativity of the core group. All three of the 2001 models of foot and mouth disease treated the farm as the host unit—effectively taking a metapopulation approach with each farm acting as a susceptible patch for the infection to colonize. This leads to the second form of host heterogeneity, the differences between farms. Primarily this variability has been attributed to the number and species of livestock within the farm, although farm geography, farm fragmentation, management practices, and livestock breed could all play a significant role. As an example, the model of Keeling *et al.* (2001b) assumes that for farm i the susceptibility (σ_i) and transmissibility (τ_i) are proportional to the number of cattle $\left(N_i^C\right)$ and sheep $\left(N_i^S\right)$ present:

$$\sigma_i = s_C N_i^C + s_S N_i^S \quad \text{and}$$
$$\tau_i = t_C N_i^C + t_S N_i^S \tag{10.8}$$

where the species-level parameters, s and t, reflect both species-level heterogeneities and species-level farming practices.

Within-farm dynamics. The standard SIR-type models (eqns 10.1 and 10.3) implicitly assume that all infected hosts (in this case farms) recover at the same rate irrespective of the time since infection—leading to exponentially distributed infectious periods, with hosts in the tail of the distribution contributing to the majority of transmission. Observations of foot and mouth disease suggest that this is not the case. Infected farms usually pass through an exposed or latent period of around 4 days before they start shedding virus; it is then around a further 5 days before signs of infection appear and control measures can be applied. Therefore, the simplest assumption (used by Keeling *et al.*, 2001b) is to model the exposed and infectious periods as discrete intervals of a constant length. A more realistic assumption is to pick the periods from gamma or normal distributions that most closely match the observed variance, including the fact that early signs of infection in sheep can often go unnoticed (Ferguson *et al.*, 2001a).

Spatial structure. The 2001 epidemic was primarily confined to four regions of the country: Cumbria, Devon, the Welsh borders, and Dumfries and Galloway; with Cumbria suffering the most cases. This spatial distribution, seeded by initial movements of infected animals, is a clear indication of the important role of localized transmission once a movement ban had been imposed. Further analysis refines this concept, showing that, after the initial spread through markets, 75% of all new cases were within 3 km of a previously reported infection, and 85% were within 5 km. Again all three models incorporate the consequences of spatial structure to some degree. Keeling *et al.* (2001b) chose to capture this effect through a localized transmission kernel, using the known location of each farm in the UK. The rate of transmission between infected farm i and susceptible farm j is modelled as:

$$\text{Rate}_{ij} = \tau_i \sigma_j K(d_{ij}) \tag{10.9}$$

where the kernel, $K(d_{ij})$, is a decreasing function of the distance between farms.

Stochasticity. Although not all models incorporated the stochastic transmission of infection, it is clear that during the initial and latter stages of the epidemic (when the number of infectious cases

were low) that random effects could play a major role. Stochasticity is therefore vitally important in predicting the end point of the epidemic; this has strategic importance, as epidemic duration is a major factor in the financial costs of an outbreak. Keeling *et al.* (2001b) therefore utilized a probabilistic approach, using the rates given above to calculate the daily probability that a given susceptible farm would be infected:

$$Prob(\text{farm } j \text{ infected}) = 1 - \exp\left(-\sum_{\text{Infectious farms, } i} \text{Rate}_{ij}\right) \qquad (10.10)$$

This term has many echoes of the simple βIS term in the SIR model (eqn 10.1).

The above description of the disease dynamics and modelling approach has clear resonance with metapopulation concepts in ecology; here farms play the role of patches or islands of various sizes and quality, with colonization between islands being a function of their separation. The only discrepancy with standard metapopulation models is the rapid culling of farms once the infection is detected.

It is important to consider why models were so successfully used during the 2001 epidemic. This stems from three basic elements. The first is the large amount of data available on the epidemic and culls, but also on the location and heterogeneity of susceptible farms. This provided a detailed set of initial conditions and the ability to accurately parameterize the necessary epidemiological processes. Secondly, the nature of transmission is relatively simple, even though the precise mechanism of some infection events is still uncertain. This was especially so after the movement ban was imposed. For human (or wildlife) diseases the movement and mixing of humans (or animals) greatly complicates the potential transmission routes—consider trying to model all human social contacts that could lead to disease spread (Ferguson *et al.*, 2005). In contrast, the sessile nature of farms simplifies the spatial dynamics, and has more in common with models of the spread of plant pathogens. Finally, the models were driven by well-posed and non-intuitive questions. Most importantly, although it is clear that the pre-emptive culling of livestock near infected farms can reduce the number of cases, is this reduction sufficient to lead to a total saving of livestock in the long term? In more precise terms, is the sum of culled and infected farms (or livestock) less than the number of infected farms (or livestock) in an uncontrolled epidemic? This is a complex nonlinear question that requires the use of accurate, well-parameterized models.

Despite the inherent differences between the three models that were used during the 2001 epidemic, remarkably similar conclusions were drawn, as follows.

1 The culling of animals on infected premises and on dangerous contacts (those farms thought to be at high risk of infection) should continue.
2 The culling on infected premises and dangerous contacts should take place as rapidly as possible and hence reduce the duration of the infectious period.
3 Further localized culling would reduce the total number of farms lost by simultaneously removing infected farms and the susceptible farms that are at the greatest risk due to their proximity to infectious sources.

The uniformity of these conclusions lies in the use of models that all agree with the observed epidemic pattern, and which agree on the initial state of the population and the effectiveness of the controls. Figure 10.6 shows the predicted epidemic profiles under a range of culling strategies; while prompt implementation is always advantageous, the complex trade-off between control of the epidemic and loss of livestock due to culling is clear.

Future epidemics of foot and mouth disease in the UK would probably also feature vaccination as a major control measure (Keeling, 2005; Keeling *et al.*, 2006). The same messages about promptness apply as for culling, although the main dynamic complexity of 'vaccinate to live' policies involves how to optimize spatial vaccination strategies rather than trade-offs between culling and slaughter (Keeling *et al.*, 2003). Analyses of the potential of vaccination for foot and mouth disease reveal particularly clearly the need for both spatially detailed stochastic epidemic models and

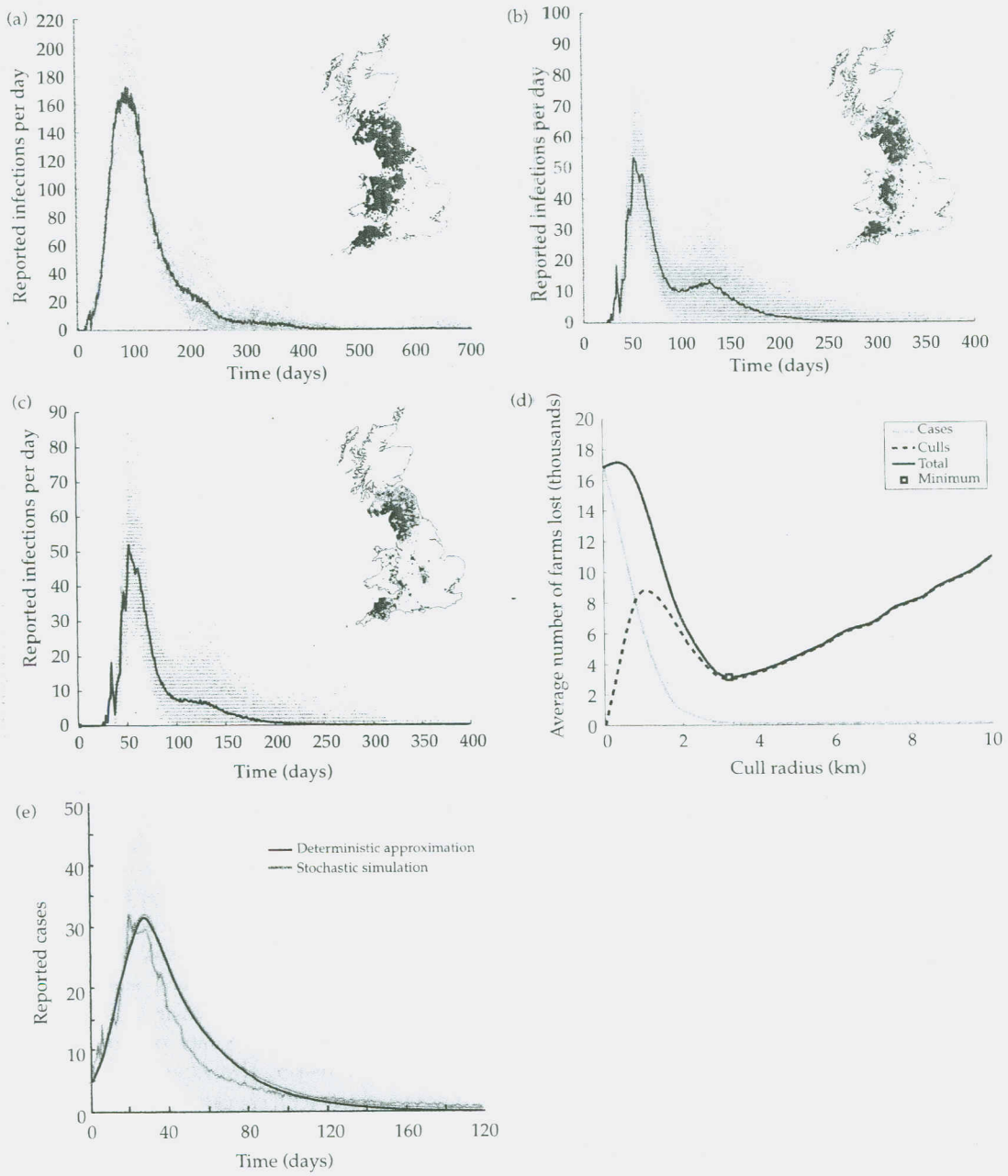

Figure 10.6 Model results from simulations of the 2001 UK outbreak of foot and mouth disease. Graphs (a)–(c) show the number of daily reported cases from 100 simulated epidemics (grey dots) together with the mean for different control options; the inset in each case shows the spatial distribution from one replicate using the same scheme as Figure 10.5b. All control options follow the 2001 temporal pattern, ramping up from low initial values and slow response times. Graph (a) is when only farms reporting infection have their livestock culled (*IP* culling), graph (b) includes the additional culling of dangerous contacts (akin to contact tracing), and graph (c) also includes culling of contiguous premises (akin to local spatial culling). Finally, graph (d) exemplifies the difficult trade-off between sufficient and overwhelming control: whereas culls within a large radius dramatically reduce the epidemic size, beyond some radius (about 3.2 km) the cull starts to be detrimental (data kindly provided by Dr Mike Tildesley). Graph (e) compares simulations of the effects of mass reactive vaccination (300 000 cattle per day), generated by repeated iterations the spatial stochastic simulation model described in the text (grey area and mean trajectory of persisting epidemics), compared with a simple core-group epidemiological model (black line); see Keeling *et al.* (2003; box 1) for full details.

more simple formulations to help interpret them. Figure 10.6e (Keeling *et al.*, 2003) compares two models from opposite ends of the spectrum between detail and simplicity. The grey bars show the range of results from 100 stochastic simulations of the full spatial model defined by eqns 10.8–10.10 with culling of infected premises and mass vaccination. The black line is the output of a simple deterministic model that has a core group of farms that are both more susceptible and more infectious. The simple and the detailed models agree on many features of the epidemic—in particular its initial rise and overall shape. However, only the stochastic simulation model is able to capture the long, low tail of the epidemic (which, in reality, dragged on to September 2001). Although small, this tail was of enormous practical importance to the UK because it determined the date from which calculations about the resumption of trade were made. Thus simple models can be invaluable for understanding how heterogeneity impacts the time course and the control of this epidemic. But for some features only a detailed model is adequate to capture observed patterns. The key to good modelling is knowing when simple models will suffice; this relies in turn on an understanding of host–pathogen natural history and a sound statistical comparison of model results with epidemiological data.

10.6 Conclusions

The study of infectious-disease dynamics has developed enormously since the 1976 and 1981 editions of *Theoretical Ecology*. A comprehensive coverage of all these developments is impossible in this short chapter. Instead, we have used a review of simple models, then more complex derivatives of them, to illustrate the place of parasitism in ecological dynamics. We argue that, in addition to their applied importance and impact on host dynamics, host–parasite relationships throw a special light on host–natural enemy interactions in general. We have focused on the synthesis of models and data for foot and mouth disease, childhood epidemics, and macroparasites in red grouse; however, a number of other examples could have illuminated general questions in ecological dynamics. Here we summarize the chapter's conclusions and point out key areas for future work.

10.6.1 Model complexity and data

Epidemiological models of specific systems are most successful when calibrated by high-quality data and a good quantitative understanding of the natural history of infection. This was the case for both foot and mouth disease and childhood infections such as measles in England and Wales. Specifically, these systems provided both an estimate of patterns of host susceptibility at epidemiologically meaningful spatio-temporal scales (farms for foot and mouth disease; cities for measles) and disease-incidence data. Together, these elements allow a key calibration: how the (imperfectly known) host-contact network determines the pattern of transmission. In contrast, recent models of smallpox and pandemic influenza (Halloran *et al.*, 2002; Ferguson *et al.*, 2005)—while often well constructed using cutting-edge modelling techniques—are necessarily hampered by a lack of epidemiological data and therefore an incomplete quantitative understanding of the precise transmission network. The public health and veterinary importance of infectious disease is likely to lead to an explosion in epidemiological and 'denominator' data collection—especially of pathogen sequence data (Kuiken, *et al.*, 2005). This should help modelling efforts; indeed models, and associated statistical developments (Keeling *et al.*, 2004), will be key for designing such sampling schemes and interpreting their results. Using models to extrapolate the fate of true emerging infections is, however, always going to be a difficult task, hedged with caveats. The use of a family of models—simple as well as complex—is key, for predicting the potential for epidemics as well as for helping to manage outbreaks once they have begun.

10.6.2 Dangerous liaisons

The intimate association of parasites and their host 'habitat' gives these interactions a special place in the ecology of host–natural enemy systems. As

discussed above, host–parasite interactions often involve and invoke heterogeneities with special dynamic consequences. Furthermore, parasites are sometimes easy to remove at the individual or local scale by treating or vaccinating their hosts. The resulting ability to manipulate the system (as in the red grouse experiments) or observe manipulations (such as the start of vaccination in childhood diseases) can enhance population-dynamic understanding. In particular, such experiments provide insights into a wider range of phase and parameter space than would normally be observed, giving clues to the dynamics away from the narrowly confined standard attractor. Such experiments (especially the implementation of control strategies) are arguably more readily performed and controlled in infectious diseases than in other ecological populations. The simple obligate interaction between some parasites and their hosts has also allowed researchers to tease apart the complex relationships between seasonality, stochasticity, and this natural-enemy interaction. Finally, the often strong effects of microparasites on host demographic rates also give them a special place in community dynamics (see also Chapter 5 in this volume). For instance, during wildlife, plant, or even human epidemics an RNA virus can cause community effects completely out of proportion to its miniscule biomass.

The importance of infectious diseases is likely to mean an ever-increasing focus on their dynamics over the next decade. As well as issues touched on above, there are a number of important areas for future development in disease dynamics. For instance, how do we characterize and model transmission network structure (Keeling and Eames, 2005; May, 2006)? How do we move beyond interactions of single hosts and parasites to model, for example, the interaction of multiple pathogens in the same host underlined by recent studies (Graham *et al.*, 2005)? However, arguably the biggest questions and opportunities arise in synthesizing population kinetics and evolutionary dynamics of host–parasite associations across scales (Nowak and May, 2000; Frank, 2002; Grenfell *et al.*, 2004). As well as their applied importance, the evolutionary dynamics of fast-evolving pathogens, such as influenza, provide a wonderful model for evolutionary ecology.