

# CHAPTER 22

## Mathematical Modeling of Infectious Diseases Dynamics

M. Choisy,<sup>1,2</sup> J.-F. Guégan,<sup>2</sup> and P. Rohani<sup>1,3</sup>

<sup>1</sup>*Institute of Ecology, University of Georgia, Athens, USA*

<sup>2</sup>*Génétique et Evolution des Maladies Infectieuses UMR CNRS-IRD, Montpellier, France*

<sup>3</sup>*Center for Tropical and Emerging Global Diseases, University of Georgia, Athens, USA*

“As a matter of fact all epidemiology, concerned as it is with variation of disease from time to time or from place to place, must be considered mathematically (. . .), if it is to be considered scientifically at all. (. . .) And the mathematical method of treatment is really nothing but the application of careful reasoning to the problems at hand.”

—Sir Ronald Ross MD, 1911

### 22.1 INTRODUCTION

The concealed and apparently unpredictable nature of infectious diseases has been a source of fear and superstition since the first ages of human civilization (see Chapters 31 and 40). The worldwide panic following the emergence of SARS and avian flu in Southeast Asia are recent examples that our feeling of dread increases with our ignorance of the disease [48]. One of the primary aims of epidemic modeling is helping to understand the spread of diseases in host populations, both in time and space. Indeed, the processes of systematically clarifying inherent model assumptions, interpreting its variables, and estimating parameters are invaluable in uncovering precisely the mechanisms giving rise to the observed patterns. The very first epidemiological model was formulated by Daniel Bernoulli in 1760 [11] with the aim of evaluating the impact of variolation on human life expectancy. However, there was a hiatus in epidemiological modeling until the beginning of the twentieth century<sup>1</sup> with the pioneering work of Hamer [32] and Ross [54] on measles and malaria, respectively. The past century has

witnessed the rapid emergence and development of a substantial theory of epidemics. In 1927, Kermack and McKendrick [41] derived the celebrated threshold theorem, which is one of the key results in epidemiology. It predicts – depending on the transmission potential of the infection – the critical fraction of susceptibles in the population that must be exceeded if an epidemic is to occur. This was followed by the classic work of Bartlett [9], who examined models and data to expose the factors that determine disease persistence in large populations. Arguably, the first landmark book on mathematical modeling of epidemiological systems was published by Bailey [8] which led in part to the recognition of the importance of modeling in public health decision making [7]. Given the diversity of infectious diseases studied since the middle of the 1950s, an impressive variety of epidemiological models have been developed. A comprehensive review of them would be both beyond the scope of the present chapter and of limited interest. Instead, here we introduce the reader to the most important notions of epidemic modeling based on the presentation of the classic models.

After presenting general notions of mathematical modeling (Section 22.2) and the nature of epidemiological data available to the modeler (Section 22.3), we detail the very basic *SIR* epidemiological model (Section 22.5). We explain

<sup>1</sup>During the nineteenth century research activity on infectious diseases was dominated by the clinical studies at the Pasteur school.

the assumptions made about the biological processes and their consequences from an epidemiological perspective. We then review more complex models that allow the study of endemic diseases (Section 22.6) and recurrent epidemics (Section 22.7). Section 22.8 then focuses on the analysis of epidemiological data and the estimation of model parameters. The chapter ends with some examples of practical uses of models for the development of public health policies (Section 22.9). Technical aspects are treated in boxes.

## 22.2 THE PHILOSOPHY OF MATHEMATICAL MODELING

Epidemiology is essentially a population biology discipline concerned with public health. As such, epidemiology is thus heavily influenced by mathematical theory. The reason is that most phenomena observed at a population level are often complex and difficult to deduce from the characteristics of an isolated individual. For example, the prevalence of a disease in a population is only indirectly connected to the course of disease in an individual. In this context, the use of mathematical models aims to unearth processes from a large-scale perspective.

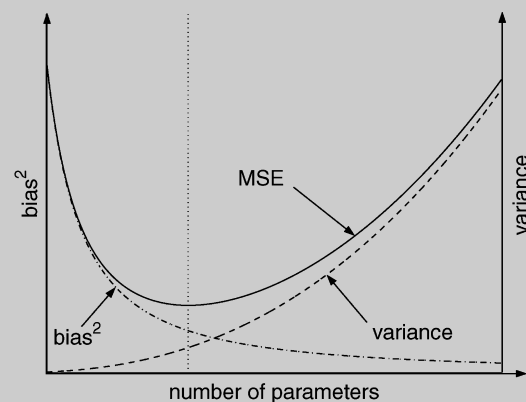
### 22.2.1 Model Complexity

A model is a caricature of reality as represented by empirical data. Models help us to understand reality because they simplify it. Consequently, all models are by definition “wrong.” There are, however, models which more closely capture essential features of reality than others – we usually refer to these as better fitting data. There is a temptation to assume that only models that are incredibly detailed (and hence “realistic”) can be useful – this is often not the case, however. A model should only be as complex as needed, depending on the questions of interest. This philosophy is referred to as Occam’s razor or the principle of parsimony and can be summarized as *the simplest explanation is the best*.

We now proceed to introduce some of the modeler’s vocabulary. A state variable is a changing quantity that characterizes the state of the system. For example, the number of infectives and susceptibles in the population are state variables of an epidemiological system. The modeler is interested in the behavior of the state variables. A parameter is a user-defined quantity that influences the value of the state variables. For example, the average duration an individual stays infectious is a parameter of an epidemiological system. The fit of a model to a data set is basically influenced by two aspects [35]. The first is related to the complexity of the model as given by the number of variables and parameters. Complicated models will usually give better fits to data than simpler models. However, simpler models are more transparent and often provide insight that is more valuable and influential in guiding thought. The choice of the optimal level of complexity obeys a trade-off between bias and variance [14] (see Box 22.1). The second aspect is related to the exact relationship between the parameters. For example, should the

### BOX 22.1 – HOW COMPLEX SHOULD A MODEL BE?

With the current power of desktop microcomputers it is tempting to build very complex models in order to fit the data the most. However, fitting the most complex model is not necessary always the best solution. Indeed, the more complex a model, the more difficult the interpretation of its outputs. Also, if a model is too complex, the modeler may not have sufficient information in the data to distinguish between the possible parameter values of the model. As said in the main text, the best-sized model depends on the purpose of the model. Given this objective, there exist quantitative methods for determining the optimal size of a model. These approaches are based on a trade-off between prediction error due to approximation (i.e., bias) which decreases as model complexity increases, and prediction error due to estimation (i.e., variance) which increases as model complexity increases as shown in the figure below [14]. The consequence is that for any model and amount of data, the total prediction error (proportional to the mean squared error) will decrease and then increase as model complexity increases, thus evidencing an optimal level of model complexity.



The mean squared error (MSE) is equal to  $MSE = \text{variance} + \text{bias}^2$ . As the number of parameters increases the bias decreases and the variances increase, defining an optimum number of parameters corresponding to the minimum of the MSE, as materialized by the vertical dotted line in the above figure.

transmission process be linear or nonlinear? Again, it is important to realize that the nature of such a relation does not need to be totally correct for the model to be useful. Modelers speak of structural stability, which refers to whether small changes in the model assumptions result in substantial changes in prediction.

### 22.2.2 Model Formulation and Hypothesis Testing

A mathematical model is a set of equations, which are the mathematical translation of hypotheses (or assumptions). When interpreting model predictions, it is thus important to bear in mind the underlying assumptions. By definition, an assumption is an unverified proposition, tentatively accepted to explain certain facts or to provide a basis for further investigation. For example, one can construct a model 1, assuming

that the probability a susceptible gets infected is proportional to the number of infectives and a model 2 assuming that this probability is independent of the number of infectives. In such an instance of competing hypotheses, the data can act as an arbitrator by telling which model is more consistent with the data [59]. In modern statistics the fit of a model to a data set is measured by its likelihood (see Box 22.2). Comparison of models is thus based on the comparison of their likelihoods. As the likelihood of a model naturally increases as the

#### BOX 22.2 – LIKELIHOOD FUNCTIONS

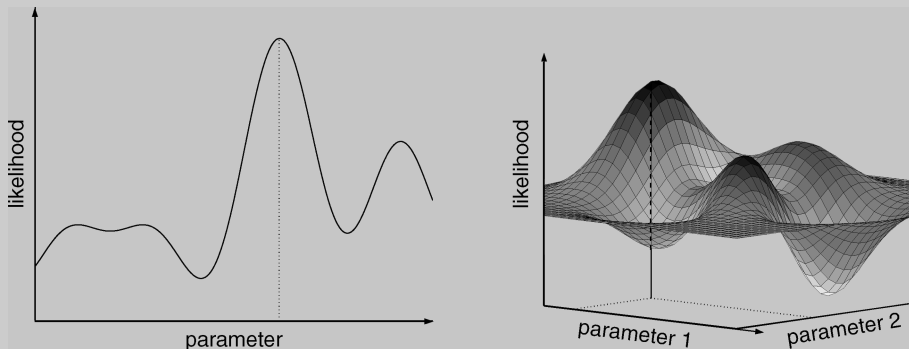
The likelihood of a model is a measure of the probability that the model is the appropriate description of the reality, given the data:  $L(\text{model} \mid \text{data}) = \text{Pr}(\text{model} \mid \text{data})$ . One powerful point of the likelihood function is that the term “model” includes not only the mean trend but also the variance, that is the distribution of the errors around the mean trend. Whereas the classical least square method implicitly assumes a normal distribution of errors, the likelihood methods allow considering any error distribution. For example, suppose that  $\mathbf{d}$  is a vector of data and  $\mathbf{m}$  a vector of model predictions with a mean trend depending on one parameter  $x$ . Assuming now that the errors are normally distributed with a variance  $\sigma^2$  then the likelihood of one prediction of the model reads

$$L(\mathbf{m}_j(x), \sigma^2 \mid \mathbf{d}_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(\mathbf{d}_j - \mathbf{m}_j)^2}{2\sigma^2}$$

If the vector of data is a time series – as often the case for epidemiological data – then the data points are not independent. However, if the noise has a large magnitude – as often the case for epidemiological data too – the approximation of independency between the data points becomes acceptable. In that case the likelihood function of the model reads

$$L(\mathbf{m}(x), \sigma^2 \mid \mathbf{d}) = \prod_i L(\mathbf{m}_i(x), \sigma^2 \mid \mathbf{d}_i) = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(\mathbf{d}_i - \mathbf{m}_i)^2}{2\sigma^2}$$

We thus end with a function which depends on two parameters  $x$  and  $\sigma^2$ . This likelihood function can be used for two different purposes. First, this function can be used to estimate parameters  $x$  and  $\sigma^2$ , good estimations of them being values that maximize the likelihood function as shown in the figures below with one and two parameters.



If the search of the maximum of the likelihood is straightforward when the function depends on one parameter, it becomes more complicated when the number of parameters increases. Microcomputers now allow the use of efficient numerical algorithm to find the maximum of such multiple dimensional surfaces. Among the most popular are the Newton and the Nelder-Mead algorithms [51]. Second, the expression of a likelihood function allows the comparison of different competing models, using either the likelihood ratio test or the Akaike information criterion (see main text).

number of its parameters increases (see above Section 22.2.1), it is necessary that the likelihood comparisons correct for the complexities of the models. There exist two major procedures for model likelihood comparison [37].

In a classic null hypothesis approach, the likelihood ratio test (LRT) is the most commonly used procedure. Two models are said nested when one is a particular case of the other. Twice the difference between log-likelihoods of two nested models follows a  $\chi^2$  distribution with degree of freedom equal to the difference between the numbers of parameters of the two models. A more complex is thus retained if its likelihood is significantly higher than the one of a simpler model, as judged from the  $\chi^2$  statistic.

We can alternatively use model selection criteria to rank any (nested or not) competitive models. These criteria are basically constructed as a likelihood value corrected for the complexity of the model. The most used of these criteria is the Akaike information criterion (AIC) [2] defined as  $AIC = 2(p-LL)$ , where  $p$  is the number of parameters and  $LL$  is the logarithm of the likelihood.

### 22.2.3 Stochastic Versus Deterministic Models

Deterministic models are those in which there is no element of chance or uncertainty. As such, they can be thought to account for the mean trend of a process only. Stochastic models, on the other hand, account not only for the mean trend but also for the variance structure around it. In an epidemiological context, there are two main kinds of stochasticity: demographic and environmental. Demographic stochasticity reflects the fact that while all individuals may be subject to the same possible events with the exact same probabilities, chance events may result in differences in the fates of individuals. When a phenomenon is the sum of a large number of small individual effects (as disease propagation in large population), the weak law of large numbers diminishes the effects of demographic stochasticity and a deterministic model becomes appropriate. In contrast, when the population is small, random events cannot be neglected and a stochastic model is necessary. Environmental stochasticity refers to the situation where there is variation in the probability associated with an event. Consequently, some parameters of stochastic models may be uncertain and characterized by a probability distribution instead of a constant value. For fixed starting values, a deterministic model will always produce the same result whereas a stochastic model will produce many different outputs, depending on the actual values the random variables take.

## 22.3 THE NATURE OF EPIDEMIOLOGICAL DATA

Epidemiology is fundamentally a data-driven discipline, and a key element in this research field is being able to link mathematical models to data. Epidemiological data are generally based on the disease notifications reported by medical doctors, veterinarians, or agronomy engineers. Epidemiologists usually consider incidences defined as the number of new

cases per unit of time and prevalences referring to the number of diseased people, ideally at one instant, and in practice over a short period of time. Incidences thus reflect the dynamics of the disease whereas prevalence is more related to the static properties of the disease. Epidemiological data may further be stratified by age, sex, social status, geographical location, and so on. In Section 22.6.2 we will see that stratification by age is of particular interest as age reflects time [7]. Moreover the survey can be carried out longitudinally (i.e., through time) or horizontally (i.e., at one instant or over a short period of time). In the first case, where the data are in the form of a time series, it is important to realize that the data of the series are not independent. Indeed, the number of new cases reported in a given week is likely to be close to cases reported during the previous week. Consequently, the statistical analysis of time series requires the use of specific tools presented in Section 22.8.2. Epidemiological data sets are often accompanied by demographic data such as the population size and the per capita birth rate in different localities and at different dates. This is of primary interest as the endemic state of an infectious disease is often dependent on host social and demographic factors.

Such data sets currently exist for a variety of diseases, in different locations and over several decades. Some of these data bases are available from the Internet, as the one used to draw Figures 22.6, 22.10, and 22.12. Other data sets can easily be requested from governmental health services. The quality of the data set is often related to its accuracy in terms of disease diagnose, spatial location, and notification frequency (weekly, monthly, or yearly).

## 22.4 CHILDHOOD MICRO-PARASITIC INFECTIONS

There exist a variety of epidemiological models and an exhaustive review of them cannot be performed in one single chapter. As a result, we will focus our attention on some of the most frequently used models in order to highlight the general approaches and the main results. We will thus be specifically interested here in childhood micro-parasitic infections. The distinction between microparasites and their counterpart macroparasites is not clear-cut and actually reflects more the way they are modeled than biological realities [7]. However, microparasites tend to refer to small-size parasites (viruses, bacteria, or protozoan) with fast and direct reproduction within the host. Childhood microparasitic diseases usually transmit by direct contact through droplets and the infectivity is generally high. The host usually recovers from the infection and acquires immunity for some time (often for life). The disease generation length (i.e., the duration between the infection and the clearance by the host immune system) is generally short relative to the host life expectancy. Because of the fast and direct reproduction within the host, it makes sense to model the dynamics of microparasitic diseases according to the host clinical status with compartmental models. We call childhood diseases those diseases which confer a lifelong

immunity. As the infectiousness of microparasitic diseases is usually high, the lifelong immunity makes the mean age at infection generally low, hence the name. Common childhood microparasitic infectious diseases include measles, rubella, chickenpox, mumps, whooping cough, and so on.

## 22.5 A SIMPLE EPIDEMIC MODEL

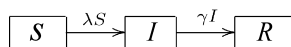
The idea behind compartmental models is to divide the host population into a set of distinct classes, according to its epidemiological status. One simple such model is the *SIR* formalism which classifies individuals as Susceptible to the disease (*S*), currently Infectious (*I*), and Recovered (*R*). The total size of the host population is then  $N = S + I + R$ . For childhood diseases there is no vertical transmission and thus individuals are born in the susceptible class (after any period of maternally derived immunity is passed). Upon contact with an infectious individual, susceptibles may get infected and move into the infectious class. Once the immune system clears the infectious agents, infecteds become immune and move to the recovered class (Figure 22.1).

### 22.5.1 Transmission Process

The transmission process is at the heart of any epidemiological model. To describe it, epidemiologists usually consider the force of infection  $\lambda$  defined as the per capita rate of acquisition of the infection. More precisely,  $\lambda(t)\Delta t$  is the probability that a given susceptible individual will acquire the infection in the small interval of time  $\Delta t$  [34].

For airborne disease, the tradition has long been to consider the force of infection proportional to the number of infectious individual:  $\lambda = \alpha I$ . There is thus an analogy with the concentration of two chemical agents to which the law of mass action applies. However, humans obviously do not behave in exactly the same way as molecules in solutions as the daily contact patterns of people are often similar in large and small communities [34, 46].

Consider instead that the average number of contacts of a person per unit time is the constant  $\beta$  combining a multitude of epidemiological, environmental, and social factors that affect transmission rates [7]. Among these contacts, the number of contact with infectives is thus  $\beta I/N$ . Assuming that contacts are sufficient for transmission, the number of new cases per unit time is then  $S\beta I/N$ . Thus, in this case  $\lambda = \beta I/N$ , instead of  $\lambda = \alpha I$ . Fits to real data have proved that



**Fig. 22.1.** A simple SIR epidemic model. The host population is divided into three compartments, according to their epidemiological status: susceptibles (*S* ind.), infectives (*I* ind.), and recovered (*R* ind.). Individuals move to the susceptible class to the infective class, to the recovered class according to the arrows.  $\lambda$  is the force of infection, that is, the probability that a susceptible individual gets infected, and  $\gamma$  is the recovery rate.

the frequency-dependant transmission process  $\lambda = \beta I/N$  is more appropriate for human airborne diseases than the density-dependant one  $\lambda = \alpha I$  [7]. The parameter  $\alpha$  has no clear epidemiological interpretation but can be related to  $\beta$  as  $\alpha = \beta/N$ . McCallum et al. [46] explored other forms of the transmission process, including nonlinear ones, and studied their influence on the epidemiological conclusions.

### 22.5.2 Between-Compartment Flux of Individuals

A common assumption is that the movements out of one compartment into the next one are governed by constant rates [7]. For each time unit a constant number of individuals leave one compartment to the next, regardless to the time they spent in their compartment. The choice of this assumption is essentially motivated by an ease of mathematical tractability in a deterministic setup with ordinary differential equations. However, the assumption of a flux of individual at a constant rate  $r$  corresponds to exponentially distributed waiting times in the compartments. The parameter of the negative exponential distribution is  $r$  and thus the mean of the distribution  $1/r$  (see Box 22.3). Analysis of real data reveals instead that each individual tends to spend a constant duration in each compartment [39, 43]. Models accounting for such realistic distributions of waiting times would imply the use of more sophisticated mathematics such as integro- or delay-differential equations. For didactic reasons we will here restrict our attention to the simplest and most used models based on simple ordinary differential equations and refer the reader interested in more realistic ones to [43] and [39]. Keeling & Grenfell [39] and Wearing et al. [60] showed that the assumptions on the waiting times can strongly influence the model outputs.

### 22.5.3 Basic Reproduction Number and Threshold Effects

One of the most fundamental quantities used by epidemiologists is certainly the basic reproduction number  $R_0$ . For microparasites it is defined as the expected number of secondary cases following the introduction of one infectious individual into a fully susceptible population [7]. We understand from here that  $R_0$  has a threshold value in the sense that a disease must have  $R_0 > 1$  to invade a host population, otherwise it disappears right after its introduction. The replacement number  $R$  is the average number of secondary infections produced by a typical infective during its entire period of infectiousness. At the introduction of one infective into a fully susceptible population  $R = R_0$  and then  $R$  decreases. At endemic equilibrium we will have, by definition,  $R = 1$  (see Sections 22.6.1 and 22.9.1.1).

### 22.5.4 Deterministic Setup and Dynamics Analysis

For large populations, deterministic models with continuous variations of population sizes provide a good description of the disease behavior. Epidemic models are used to describe

**BOX 22.3 – MODELING THE INFECTIOUS PERIOD**

In the classic *SIR* model it is usually assumed that the individuals leave the infectious class at a constant rate. Even if this assumption seems the most intuitive, it is not always the most realistic in terms of the duration individuals stay infective. In this box we detail the consequences of the constant recovery rate assumption on the distribution of the infective periods, and propose an alternative which yields more realistic distributions [39,43,60].

Our random variable is the time of recovery since the infection. For discrete random variables (e.g., number of individuals) it is easy to define a probability distribution  $\Pr\{Z = k\} = f_k$  (as in Section 22.5.5) and then to define a cumulative distribution function  $F(z) = \Pr\{Z \leq z\}$ . For continuous variables, like here, the time of recovery since infection, it is impossible to define a probability of each time as there is an infinity of such times. The approach is then to first define a cumulative distribution and then express a probability density function from this cumulative distribution. The idea is to consider the probability associated with a short interval  $\Delta z$  of the random variable  $z$ .

$$\begin{aligned} \Pr\{z \leq Z \leq z + \Delta z\} &= F(z + \Delta z) - F(z) \\ &= F'(z)\Delta z + o(\Delta z) \end{aligned}$$

The derivative  $F'(z)$  of the cumulative distribution  $F(z)$  is by definition the probability density function.

Let us now apply this method to the time of recovery since the infection. As done in Section 22.6.3, we can express the probability of an infective to recover in the time interval  $\Delta t$  as

$$\Pr\{\text{recovery in } (t, t + \Delta t] \mid \text{no recovery in } (0, t]\} = \gamma\Delta t + o(\Delta t)$$

where  $t$  is the time since infection and  $\gamma$  is a fixed constant. The cumulative distribution is defined as  $F(t) = \Pr\{\text{no recovery in } (0, t]\}$ . For an infective not to recover in the interval  $(0, t + \Delta t]$ , he must first not recover in the interval  $(0, t]$  and then not recover in the next  $\Delta t$ . Assuming that these events are independent gives

$$\begin{aligned} F(t + \Delta t) &= F(t)[1 - \gamma\Delta t + o(\Delta t)] \\ \Leftrightarrow \frac{F(t + \Delta t) - F(t)}{\Delta t} &= -\gamma F(t) + \frac{o(\Delta t)}{\Delta t} \end{aligned}$$

Taking the limit as  $\Delta t \rightarrow 0$  gives

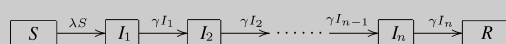
$$\frac{dF}{dt} = -\gamma F(t)$$

which, after integration and setting  $F(0) = 1$  (i.e., no recovery before the infection), yields

$$F(t) = e^{-\gamma t}$$

Thus, infectious periods are exponentially distributed with a mean infectious duration equal to  $1/\gamma$  (see dashed curve on the figure below). Inspecting real data, it seems that the infectious period does not follow an exponential distribution but rather seems to be of constant duration. To account for such more realistic distributions, we need to relax the assumption that the probability of recovery does not depend on the time since infection. There are several ways to do that, including integro-differential and partial differential formulations, but the simplest one is certainly the method of stages.

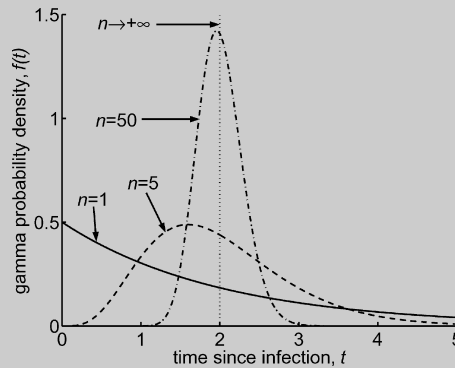
The basic idea of the method of stages is to replace the infective compartment by a series of  $n$  successive infective compartments, each with an exponential distribution of the same parameter:



The total duration of the infectious period is thus the sum of  $n$  identical and independent exponential distributions, which leads to a gamma distribution of the infectious durations:

$$f(t) = \frac{(\gamma n)^n}{\Gamma(n)} t^{n-1} e^{-\gamma n t}$$

where  $\Gamma(n)$  is the gamma function. The variance of such a distribution is  $1/(n\gamma^2)$ . Notice that when  $n = 1$  we find back the above-presented exponential distribution, when  $n$  gets large, the gamma distribution tends toward a normal one and when  $n \rightarrow \infty$  we have the delta (fixed duration with no variance) distribution (see figure below).



The above figure shows gamma distribution for  $\gamma = 0.5$  and various values of the number  $n$  of classes. When  $n = 1$  we have the exponential distribution and when  $n$  increases the distribution tends toward a normal one. Ultimately, when  $n \rightarrow \infty$  the gamma distribution converge toward the delta distribution with a variance equal to zero. Note that for all values of  $n$  the mean is equal to  $1/\gamma = 2$ .

rapid outbreaks that occur in very short periods of time, during which the host population can be assumed to be in a constant state [18, 19]. A mathematical description of the fluxes of individuals of Figure 22.1 is given by the following set of differential equations:

$$\frac{dS}{dt} = \beta S \frac{I}{N} \quad S(0) = S_0 \geq 0 \quad (22.1)$$

$$\frac{dI}{dt} = \beta S \frac{I}{N} - \gamma I \quad I(0) = I_0 \geq 0 \quad (22.2)$$

$$\frac{dR}{dt} = -\gamma I \quad R(0) = R_0 \geq 0 \quad (22.3)$$

where  $\gamma$  is the recovery rate. Since the duration of the epidemic is short, this model has no host vital rate. In consequence, the total host population size  $N = S + I + R$  is constant and only two of the above equations are necessary to totally account for the disease behavior. Dividing the first two Equations (22.1) and (22.2) by the constant host population size  $N$  yields

$$\frac{ds}{dt} = -\beta s i \quad s(0) = s_0 \geq 0 \quad (22.4)$$

$$\frac{di}{dt} = \beta i s - \gamma i \quad i(0) = i_0 \geq 0 \quad (22.5)$$

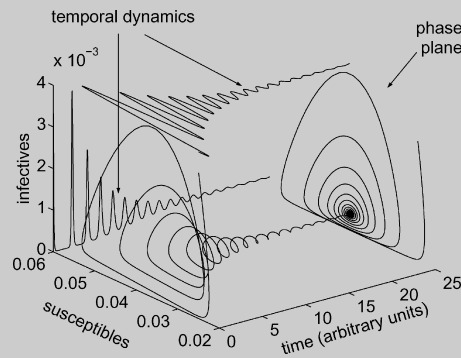
where  $s(t) = S(t)/N$  and  $i(t) = I(t)/N$ . The basic reproduction number then reads  $R_0 = s_0 \beta/\gamma$ . Thus we can express the threshold on  $R_0$  as follows. When  $s_0 < \gamma/\beta$ , on average each infective produces less than one infective and thus the number of infectives diminishes to reach 0 as time passes on. When  $s_0 > \gamma/\beta$ , the number of infectives first increases to then decrease toward 0, producing this characteristic epidemic peak. This threshold effect is illustrated on the phase plane (see Box 22.4) of Figure 22.2. We can see on this figure that when  $s_0 < \gamma/\beta$  the proportion of infectives decreases toward 0, and when  $s_0 > \gamma/\beta$  the proportion of infectives first increases to then decrease toward zero. In any case the proportion of infectives ends at zero whereas the ultimate value  $s_\infty$  of the proportion of susceptibles depends on the initial proportions  $s_0$  and  $i_0$  of susceptibles and

**BOX 22.4 – GRAPHICAL TOOLS TO STUDY DYNAMICAL SYSTEMS**

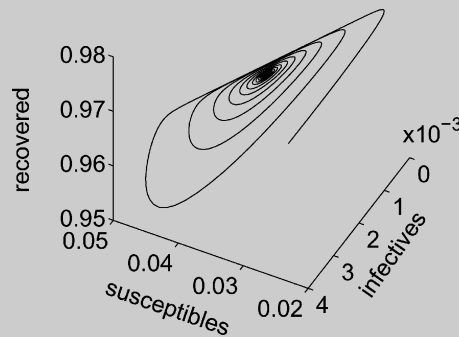
In this box we present two graphical tools facilitating the study of dynamical systems. The first one is the phase plane. Consider for example the endemo-epidemic *SIR* model of the main text

$$\begin{aligned} ds / dt &= \mu - \beta is - \mu s & s(0) &= s_0 \geq 0 \\ di / dt &= \beta is - \gamma i - \mu i & i(0) &= i_0 \geq 0 \end{aligned}$$

We can solve this system and draw the temporal dynamics of each of the state variable,  $s(t)$ ,  $i(t)$ , and  $r(t) = 1 - s(t) - i(t)$ . A phase plane plots the behavior of one state variable as a function of another state variable. Temporal dynamics and phase plane are thus two different ways of visualizing the same reality as exemplified on the figure below.



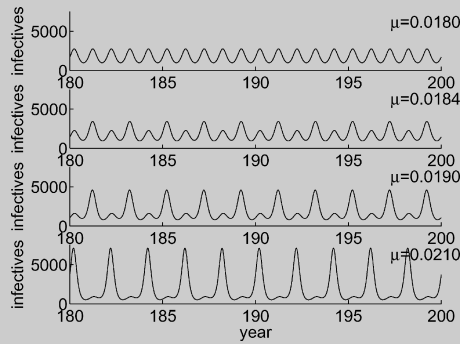
The phase plane of the above figure is the same as the one of Figure 22.7. There is no time dimension on phase planes but the trajectory of the dynamics is usually indicated by arrows (see Figures 22.2, 22.5, and 22.7). A phase plane can also be drawn for three state variables like on the figure below.



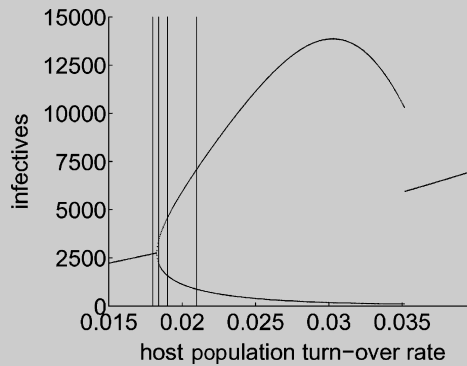
The second tool is relative to the complexity of a dynamics. Consider the same *SIR* model but now with a varying contact rate to sustain the oscillations (see Section 22.7.1):

$$\beta(t) = \beta_0(1 + \beta_1 \cos(2\pi t)) \quad 0 \leq \beta_1 \leq 1$$

Running this model with different values of the host population turnover rate  $\mu$  yields qualitatively different disease dynamics: on the figure below the dynamics changes from annual to bi-annual when  $\mu$  increases from 0.0180 to 0.0210.



These qualitative changes on the dynamics are called bifurcations and the parameter we explore the influence (here  $\mu$ ) is called the control parameter. Bifurcation diagrams allow to visualize the effect of a control parameter on the complexity of the dynamics. For each value of the control parameter the simulated dynamics is sampled at regular time intervals. Imagine, for example, that we sample the dynamics every year. Then, an annual dynamics will give one point on the bifurcation diagram (each year the dynamics recovers the same value), whereas a biannual dynamics will give two points (one for the odd years and the other for the even years).



The above figure shows the bifurcation diagram of the disease dynamics with  $\mu$  as the control parameter. The four vertical lines materialize the  $\mu$  values corresponding to the above four time series. This diagram predicts that the disease dynamics is biannual for  $\mu$  between 0.0183 and 0.0351 and annual for  $\mu$  between 0.0150 and 0.0183 and between 0.0351 and 0.0400. At  $\mu = 0.0183$  the switch from annual to biannual is progressive whereas at  $\mu = 0.0351$  the switch from biannual to annual is sharp. Between 0.0183 and 0.0351 the disease oscillations reach their maximum at  $\mu = 0.0303$ .

infectives, respectively, as expressed by the following implicit equation [19]:

$$i_0 + s_0 - s_\infty + \log(s_\infty / s_0) / \sigma = 0 \tag{22.6}$$

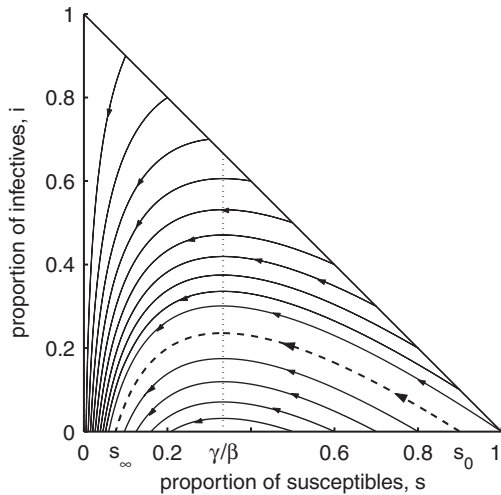
We can see from Figure 22.2 that the higher the initial proportion of susceptibles  $s_0$ , the lower the proportion of individuals who do not get diseased during the epidemic. This is known as overshoot phenomenon [19].

### 22.5.5 Stochastic Dynamics and Probability of an Epidemic in a Small Population

The deterministic *SIR* model presented above highlights a threshold value on the basic reproduction number with an epidemic when  $R_0 > 1$  and no epidemic when  $R_0 < 1$ .

However, observations on real data reveal that  $R_0 > 1$  does not guarantee an epidemic in the population [9]. The cause of this discrepancy between model prediction and observed data is that the deterministic *SIR* model is a good approximation of the epidemic dynamics only when dealing with large populations (see above Section 22.2.3), which is clearly not the case when we are interested in the initial epidemic growth following the introduction of one infective into a fully susceptible population. As the initial number of infectives during the initial epidemic growth is by definition very small, demographic stochasticity may play an important role in the start of an epidemic. The theory of branching processes is a useful framework to derive the probability that an epidemic starts [19].

Consider that the number of people infected by one infective follows a given probability distribution  $\{q_k\}_{k=0}^\infty$ . Thus,



**Fig. 22.2.** Phase plane of the *SIR* epidemic model of Equations (22.4) and (22.5). The arrowed lines show the trajectories of the dynamics. The dashed arrowed line shows one particular trajectory with its initial and final proportions of susceptible  $s_0$  and  $s_\infty$ , respectively. The vertical dotted line is the threshold  $\gamma/\beta$  on the value of the proportion of susceptibles (see main text). There is an epidemic only when  $s_0$  is above this threshold.

any infective infects  $k$  individuals with the probability  $qk$  and  $\sum_{k=0}^{\infty} q_k = 1$ . The basic reproduction ratio  $R_0$  can then be expressed simply as the expected number of individuals infected by one infective:  $R_0 = \sum_{k=1}^{\infty} kq_k$ . Then, we need to introduce the reader to one fundamental tool of branching processes: the generation function defined as

$$g(z) = \sum_{k=0}^{\infty} q_k z^k \quad 0 \leq z \leq 1 \quad (22.7)$$

Among the interesting properties of the generating function are  $g(0) = 0, g(1) = 1$ , and  $g'(1) = R_0$  [19]. Let  $z_n$  be the probability that the disease disappears from the population after  $n$  generations of transmission events. It can be shown that  $z_n = g(z_{n-1})$  [33]. As the function  $g$  is increasing, the sequence  $z_n$

is increasing and tends toward a limit  $z_\infty$ . By definition  $z_\infty$  is the probability that the disease introduced by one individual into a fully susceptible population will go extinct. Thus  $z_\infty$  is the solution of the equation  $z = g(z)$ . It can be shown that  $z_\infty = 1$  for  $R_0 \leq 1$  and  $0 < z_\infty < 1$  for  $R_0 > 1$  [19]. Depending on the exact form of the infectious process, the solution  $z_\infty$  of the equation  $z = g(z)$  can not always be expressed explicitly. For example, assuming that the number of infections during a constant time interval is according to a Poisson process, we end up with the following implicit expression of  $z_\infty$  [19]:

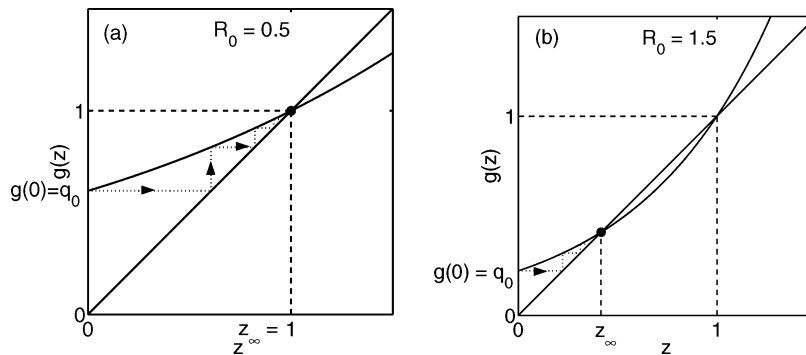
$$z = z \exp(R_0(z - 1)) \quad (22.8)$$

which can be easily solved graphically (see Figure 22.3).

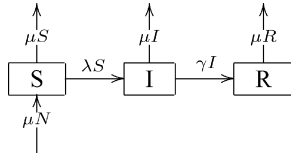
## 22.6 A SIMPLE ENDEMIC MODEL

### 22.6.1 Deterministic Dynamics

Epidemic models presented in the above section are used to describe rapid outbreaks that occur in very short period of time, during which the host population can be assumed to be in a constant state. Such models thus do not need to account for the host population dynamics as governed by births and deaths. On longer period of times individuals will die and births will feed the population with new susceptibles, possibly allowing the disease to persist in the population at a low and constant prevalence. We then say that the disease is in an endemic state in the population [7]. If we are to study the endemic state of a disease, we need to construct a model that accounts for the birth and death rate of the host population. In the case of a nonfatal disease like most childhood ones in developing countries, a good approximation is to consider that the host population size  $N = S + I + R$  is constant. The dynamics of the disease can then be described by the following differential



**Fig. 22.3.** Graphical resolution of the implicit Equation (22.8). Solutions of Equation (22.8) are the intersections between the first bissectrice and the curve which are respectively the l.h.s. and the r.h.s. of Equation (22.8), in the domain of definition  $[0,1]$ .  $z_\infty = 1$  when  $R_0 < 1$  (a) and  $0 < z_\infty < 1$  when  $R_0 > 1$  (b), (see main text).



**Fig. 22.4.** A simple SIR endemic model. Same as Figure 22.1 except that now deaths remove individual from each compartment at a constant rate  $\mu$ . Also, births feed the susceptible compartment with new individuals at the same rate  $\mu$ . As the birth and death rates are equal, the total size  $N$  of the whole population remains constant, see main text.

equations which correspond to the flow diagram of Figure 22.4.

$$\frac{ds}{dt} = \mu - \beta is - \mu s \quad s(0) = s_0 \geq 0 \quad (22.9)$$

$$\frac{di}{dt} = \beta is - \gamma i - \mu i \quad i(0) = i_0 \geq 0 \quad (22.10)$$

where  $\mu$  is the host population turnover rate, that is, the birth rate equal to the death rate. Again, for the simplicity of the mathematical analysis, assume that this rate has a constant value. As explained above and in Box 22.3, the consequence of this assumption is that the age distribution follows a negative exponential distribution. The mean of this distribution (i.e., the host life expectancy  $L$ ) is equal to  $L = 1/\mu$ . This is a rather good approximation for the developing countries where the harshness of the environment imposes a similar death pressure on all the age classes [7]. However, in western countries, medical care allows most of the people to reach the natural age limit, yielding this characteristics square shape age distribution. Nevertheless, the exact form of the age pyramid does not have substantial influence on the dynamics of the disease [7].

The basic reproduction ratio now reads  $R_0 = \beta/(\gamma + \mu)$ . By definition, at equilibrium the system is in a constant state.

Thus the differentials of Equations (22.9) and (22.10) should be equated to 0:  $ds/dt = di/dt = 0$ , which yields the following system of equations:

$$\mu - \beta is - \mu s = 0 \quad (22.11)$$

$$\beta is - \gamma i - \mu i = 0 \quad (22.12)$$

Solving this system produces two equilibrium points: (i) the disease-free scenario  $(s_1^*, i_1^*, r_1^*) = (1, 0, 0)$  and (ii) the endemic case  $(s_2^*, i_2^*, r_2^*) = (1/R_0, \mu(R_0 - 1)/\beta, 1 - s_2^* - i_2^*)$ . The stability of these two equilibria depends solely on the value of the basic reproduction number, and not on the initial values of the proportions of susceptibles and infectives as in the above epidemic model. If  $R_0$  is less than unity, then the disease-free equilibrium is stable [19] and the phase plane of Figure 22.5 shows that the proportion of susceptibles increases toward 1 whereas the proportion of infectives decreases toward 0. When  $R_0 > 1$  means that the endemic equilibrium is stable [19] and Figure 22.5 shows that the proportions of susceptibles and infectives produce damped oscillations that converge toward their endemic values  $s^*$  and  $i^*$ . Linear stability analysis (see Box 22.5) reveals the natural period  $T$  and the damping time  $D$  of these damped oscillations to be approximated by

$$\hat{T} = 2\pi\sqrt{AG} \quad (22.13)$$

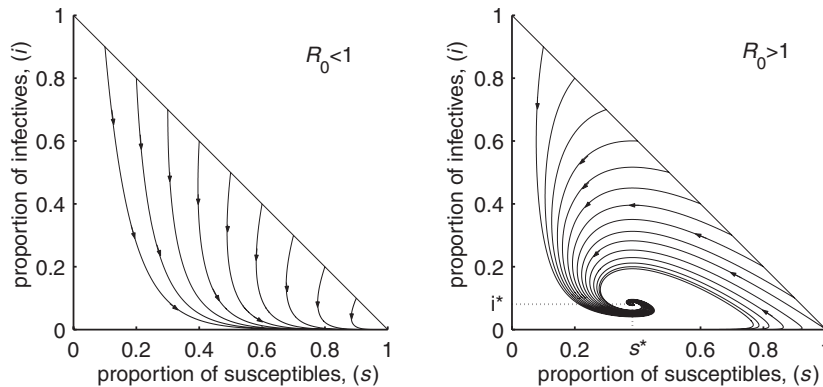
and

$$\hat{D} = 2A \quad (22.14)$$

respectively, where  $A$  represents the mean age at infection,  $A = 1/\mu(R_0 - 1)$  (see below Section 22.6.2), and  $G$  gives the ecological generation length of the infection, that is, the sum of the latent and infectious periods,  $G = 1/(\mu + \gamma)$  [7, 53].

### 22.6.2 Statics and the Average Age at Infection

Once the endemic equilibrium is reached we may be interested in the statics of the disease such as the mean age at infection. This is of importance as, first, numerous diseases are



**Fig. 22.5.** Phase plane of the SIR endemic model of Equations (22.9) and (22.10). The arrowed lines show the trajectories of the dynamics. When the basic reproduction number  $R_0 < 1$  the dynamics converges toward the stable disease-free equilibrium (left). When  $R_0 > 1$  the disease dynamics converges toward endemic equilibrium  $(i^*, s^*)$  (right).

**BOX 22.5 – LINEAR STABILITY ANALYSIS BASED ON EIGENVALUES**

Linearization approximation is a standard phase plane technique used to analyze system dynamics [42]. For an *SIR* system with a constant host population size we have the following system of two independent nonlinear differential equations:

$$\frac{ds}{dt} = \mu - \beta is - \mu s \quad s(0) = s_0 \geq 0$$

$$\frac{di}{dt} = \beta is - \gamma i - \mu i \quad i(0) = i_0 \geq 0$$

As found in the main text, the endemic equilibrium of this system is  $(s^*, i^*) = (1/R_0, \mu(R_0 - 1)/\beta)$ . Close to the endemic equilibrium, the above system can then be rewritten into the following form:

$$s(t) = s^* + \xi(t)$$

$$i(t) = i^* + \zeta(t)$$

where  $\xi(t)$  and  $\zeta(t)$  are the deviations from the equilibrium. In order to study the stability of the equilibrium, we then need to focus on the dynamics of the deviations  $\xi(t)$  and  $\zeta(t)$  [42]. Combining the above two systems, developing, and keeping only the terms which are linear in  $\xi$  and  $\zeta$ , we get

$$\frac{d\xi}{dt} = -(\beta i^* + \mu)\xi - \beta s^* \zeta + \text{NL}(\xi, \zeta)$$

$$\frac{d\zeta}{dt} = \beta i^* \zeta + \text{NL}(\xi, \zeta)$$

where  $\text{NL}(\xi, \zeta)$  contains all the nonlinear terms in  $\xi$  and  $\zeta$ . Replacing  $s^*$  and  $i^*$  by their value gives

$$\frac{d\xi}{dt} = -[\mu(R_0 - 1) + \mu]\xi - \frac{\beta}{R_0}\zeta + \text{NL}(\xi, \zeta)$$

$$\frac{d\zeta}{dt} = \mu(R_0 - 1)\xi + \text{NL}(\xi, \zeta)$$

Written in matrix form, the above system becomes

$$\begin{matrix} d\xi / dt \\ d\zeta / dt \end{matrix} = \underbrace{\begin{bmatrix} -\mu(R_0 - 1) - \mu & -\beta / R_0 \\ \mu(R_0 - 1) & 0 \end{bmatrix}}_{\mathbf{J}} \cdot \begin{bmatrix} d\xi \\ d\zeta \end{bmatrix} + \text{NL}(\xi, \zeta)$$

The Jacobian matrix  $\mathbf{J}$  is called the community matrix in ecology and its eigenvalues are indicative of the dynamics of the system [42]. The eigenvalues of the community matrix are solutions of the characteristic equation

$$\Lambda^2 - \text{Tr}(\mathbf{J})\Lambda + \det(\mathbf{J}) = 0$$

where  $\text{Tr}$  and  $\det$  refer to the trace and the determinant of the matrix, respectively. Replacing the trace and determinant by their values gives

$$\Lambda^2 - \mu R_0 \Lambda + \mu(\mu + \gamma)(R_0 - 1) = 0$$

With the approximation pertaining to the fact that  $\gamma \gg \mu$ , we end with

$$\Lambda \approx \frac{1}{2A} \pm j \frac{1}{\sqrt{AG}}$$

where  $A$  represents the mean age at infection,  $A \approx 1/\mu R_0$ , and  $G$  gives the ecological generation length of the infection, that is, the sum of the latent and infectious periods,  $G = 1/(\mu + \gamma) \approx 1/\gamma$  [7,53]. The system oscillates with a period equal to  $2\pi$  times the inverse of the imaginary part of the eigenvalue,  $\hat{T} = 2\pi\sqrt{AG}$ , and a damping time equal to the inverse of the real part of the eigenvalue,  $\hat{D} = 2A$  [42].

in endemic state in human populations and, second, the study of static properties of a disease allows the estimation of key epidemiologic parameter without requiring the long series of longitudinal notifications, often difficult to obtain in practice. The idea behind studies on the statics of diseases is that the age of the individuals reflects, in some way, time [7]. What we simply need here is horizontal data stratified by age.

Considering the age as a continuous variable, the mean age at infection is simply expressed as [7]

$$A \equiv \int_0^\infty a \frac{\lambda s(a)}{\int_0^\infty \lambda s(a) da} da \quad (22.15)$$

which is the integral sum of the age values  $a$ , weighted by the proportion of infectives of age  $a$ . Calculating this integral for a constant host population turnover rate  $\mu$  yields the intuitive relationship  $A = 1/(\lambda + \mu)$ . This means that the higher the force of infection (i.e., the probability that a susceptible gets infected), the lower the mean age at infection. Moreover, recall from above (Section 22.6.1) that  $i^* = \mu(R_0 - 1)/\beta$ . Thus,  $\lambda = \beta i^* = \mu(R_0 - 1)$ . Rearranging this equation we get  $\lambda + \mu = \mu R_0$ . An expression of the mean age at infection then becomes  $A = 1/(\mu R_0)$ . This last expression allows estimating the basic reproduction number  $R_0$  in a rather simple way as

$$R_0 = \frac{L}{A} \quad (22.16)$$

where  $L = 1/\mu$  is the host life expectancy (see Section 22.6.1).

**22.6.3 Stochastic Dynamics and Disease Persistence**

The above study of the deterministic dynamics of diseases has revealed a threshold on the value of the basic reproduction number. The disease immediately disappears after

its introduction as soon as  $R_0 < 1$  and persists at an endemic level in the host population when  $R_0 \geq 1$ . However, by inspection of real data, it appears that the condition  $R_0 \geq 1$  does not guarantee the disease persistence [9]. As already mentioned about the epidemic model (see Section 22.6.3), such persistence is dependent on the magnitude of the stochastic fluctuations around the endemic equilibrium.

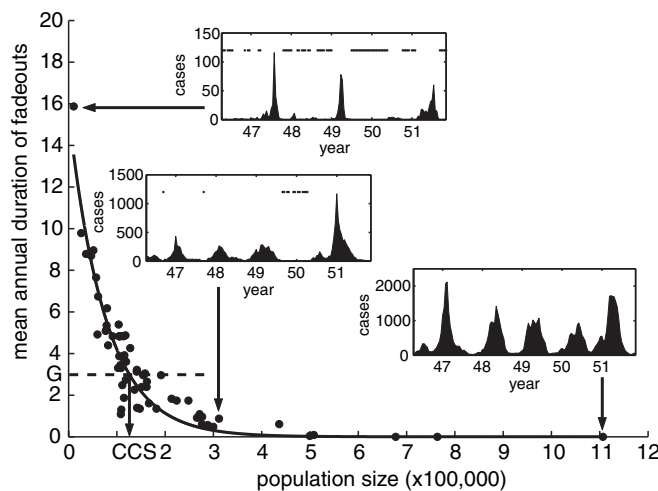
In a metapopulation context, the probability of disease extinction in one subpopulation depends on both the size of the subpopulation and the fluxes of infectives from neighbor subpopulations. Bartlett [9] has thus evidenced that there is a community size above which the disease can be maintained in population by itself and below which the disease cannot persist in the population without regular fluxes of infectives from neighbor populations. The determination of this critical community size is performed empirically by plotting the mean annual duration of periods with no cases against the size of the subpopulation. By definition, we have a period of disease fade-out if the duration of the disease extinction is longer than the disease generation length [9]. Figure 22.6 shows an example for measles in 59 cities of England and Wales in the pre-vaccine era (1944–1966). For this disease the generation length is around 3 weeks and the critical community size is estimated here at about 115,000 individuals. The critical community size is thus a quantity very easily calculated from disease notifications and which gives a good idea of the population size required for disease persistence. Intuitively we expect that the more contagious

a disease, the lower the critical community size. This explains why highly contagious diseases cannot persist in small isolated communities such as island populations or primitive Amazonian tribes.

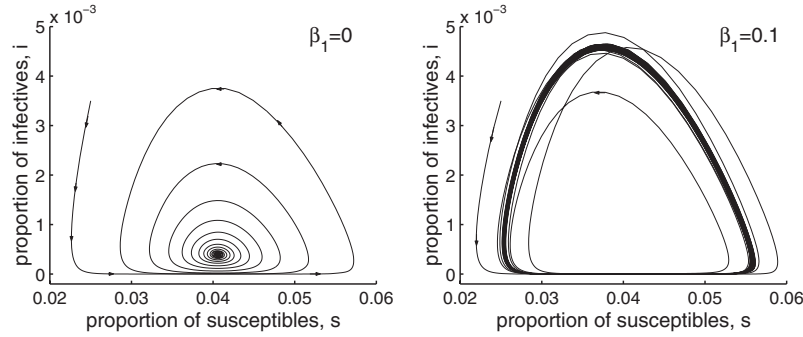
## 22.7 ENDEMO-EPIDEMIC MODELS

So far we have seen simple models that allow the study of one isolated epidemic (Section 22.5) or of diseases in an endemic state (Section 22.6). However, it appears that numerous diseases are characterized by an endemic background with regular epidemics as visible on subplots of Figure 22.6 or on Figure 22.10. We say that these diseases are in an endemo-epidemic state in the population. In this section we propose some complications of the basic endemic model that allow producing recurrent outbreaks as observed on many longitudinal surveys.

In Section 22.6.1 we have shown that the endemic model exhibits damped oscillations which converge toward an endemic equilibrium. Linear stability analysis further revealed the natural period and the damping time of these oscillations to be approximated by  $T = 2\pi \sqrt{AG}$  and  $D = 2A$ , respectively, where  $A$  and  $G$  are the mean age at infection and the disease generation length respectively (see Equations (22.13) and (22.14)). Importantly, for most epidemiologically reasonable parameter values, the damping time is typically much longer than the natural period:  $2A/T \gg 1$ . This renders the endemic equilibrium weakly stable, with relatively small per-



**Fig. 22.6.** Mean annual duration of fade-outs (i.e., local extinction) of measles against population size for 59 cities in England and Wales in the pre-vaccine era (1944–1966). Subplots show portions of time series illustrating the three levels of persistence identified by Bartlett [9]. Type I dynamics (bottom subplot: Birmingham, population of 1.1 million ind.) are regular, endemic, with no fade-out. Type II dynamics (middle subplot: Nottingham, population of 300,000 ind.) are regular but with some fade-outs (represented by black dots) in the troughs. Type III dynamics (top subplot: Teignmouth, population of 11,000 ind.) are irregular with long fade-out between the epidemics. The curve is the nonlinear regression ( $y \approx 16 \times \exp[-10^{-5} x]$ ) and its intersection with the disease generation length ( $G$ , represented by the horizontal dotted line) gives the critical community size (CCS) of the disease of around 115,000 ind. Data downloaded from <http://www.zoo.cam.ac.uk/zoostaff/grenfell/measles.htm> [27].



**Fig. 22.7.** Phase plane of the *SIR* endemo-epidemic model with a contact rate varying according to Equation (22.17) and  $R_0 > 1$ . When  $\beta_1 = 0$ , the dynamics converges toward an endemic equilibrium point (left). When  $\beta_1 = 0.1$ , the equilibrium point is destabilized and the dynamics produces sustained oscillations (right).

turbations (intrinsic or extrinsic) exciting and sustaining the inherent oscillation behavior [31]. Alternative mechanisms for this phenomenon have been proposed in the literature and all are based on the inclusion of some heterogeneity in the endemic model. Heterogeneity can be added temporally on the coefficient of transmission, spatially in the context of metapopulations, or by cohorts for age-structured models. Lastly, heterogeneity can be added statistically for full stochastic versions of the endemic model.

### 22.7.1 Varying Contact Rate

Temporal heterogeneity in the transmission rate was first proposed by Soper [56] who was attempting to explain the high amplitude outbreaks of measles in Glasgow. He demonstrated that transmission rates were high in October and declined through the academic year, with a trough in the summer months. This temporal variability in transmission rates, he argued, may be due to the considerably higher transmission rates when children are in school. Soper’s conclusions were supported by later analyses of measles, chickenpox, and mumps in some US cities, as well as measles in England & Wales [26,44]. There has been a variety of mathematical forms proposed for taking into account seasonality in the coefficient of transmission [20, 22, 24, 40]. Certainly, the most realistic take the form of a binary function, with two different values of the coefficient of transmission – one for the school terms and one for the holidays. This necessitates the knowledge of the school holidays calendar which is not always possible, particularly for historical data. A simpler form of the coefficient of variation would simply take the form of a sinusoidal wave:

$$\beta(t) = \beta_0(1 + \beta_1 \cos(2\pi t)) \quad 0 \leq \beta_1 \leq 1 \quad (22.17)$$

where the strength of seasonality  $\beta_1$  measures the amplitude of the oscillations around the baseline coefficient of transmission  $\beta_0$ . Although less realistic, this form of the coefficient of transmission produces results which are qualitatively very close to the ones

obtained with a coefficient of transmission in plateau [22]. Figure 22.7 shows that even small strengths of seasonality are able to produce sustained oscillations.

### 22.7.2 Age-Structured Models

When the time is considered as a continuous variable, the most general form of the force of infection is actually the following [7]:

$$\lambda(a, t) = \int_0^\infty \beta(a_s, a_i, t) i(a_i, t) da_i \quad (22.18)$$

where  $\beta(a_s, a_i, t)$  is the coefficient of transmission between a susceptible of age  $a_s$  and an infective of age  $a_i$  at time  $t$ . In Sections 22.5 and 22.6 we averaged this relation over both time and ages. In the above Section 22.7.1 we averaged over ages only and we defined the coefficient of transmission as a function of time (see Equation (22.17)):

$$\bar{\lambda}(t) = \bar{\beta}(t) \bar{i}(t) \quad (22.19)$$

where bars refer to age average. In the present section we average the relation 22.18 over time only. We thus have to define a coefficient of transmission as a function of age. Contrary to time, it does not really make biological sense to consider age as a continuous variable as human populations are usually aggregated by cohorts defined as the primary school children, the intermediate and high school teenagers, the college young adults, and the adults [7]. When averaging over time and considering the age variable as a discrete variable, Equation (22.18) becomes:

$$\hat{\lambda}_i = \sum_{j=1}^n \hat{\beta}_{i,j} \hat{i}_j \quad (22.20)$$

where hats refer to time average and  $n$  is the number of distinct cohorts. We thus need to define a matrix of transmission  $[\hat{\beta}_{i,j}]$ .

**TABLE 22.1.** Transition Events, and Corresponding Rates, for a Simple Stochastic SIR Model

Type of transition event	Rate	Event
1 $S \rightarrow S+1, I \rightarrow I, S \rightarrow R$	$r_1 = \mu N$	Birth
2 $S \rightarrow S-1, I \rightarrow I, S \rightarrow R$	$r_2 = \mu S$	Death
3 $S \rightarrow S, I \rightarrow I-1, S \rightarrow R$	$r_3 = \mu I$	Death
4 $S \rightarrow S, I \rightarrow I, S \rightarrow S-1$	$r_4 = \mu R$	Death
5 $S \rightarrow S-1, I \rightarrow I+1, R \rightarrow R$	$r_5 = \beta IS/N$	Infection
6 $S \rightarrow S, I \rightarrow I-1, R \rightarrow R+1$	$r_6 = \gamma I$	Recovery
7 $S \rightarrow S, I \rightarrow I+1, R \rightarrow R$	$r_7 = \delta$	Immigration of infectives

### 22.7.3 Spatially Structured Models

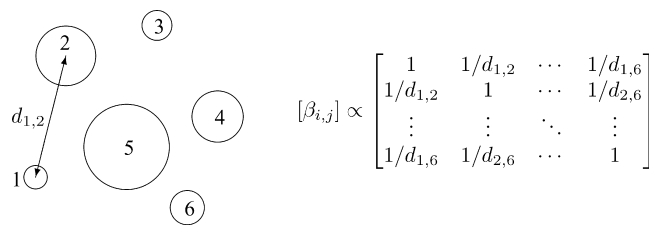
The organization of human populations in distinct cities interconnected by fluxes of individuals makes the theory of metapopulations an appropriate framework to study the spatial dynamics of infectious diseases [28, 29]. In this context the definition of a spatially structured model is pretty close to an age-structured one. We thus need to define a matrix of transmission. Two common assumptions are that this matrix is symmetric and the values  $\beta_{i,j}$  are related to the geographic distance between the cities  $i$  and  $j$  (Figure 22.8). However, given the speed of communication networks at a regional scale, a simple and widely used approximation of this metapopulation model is the island model in which all the subpopulations are linked the ones to the others by the same coupling coefficient  $\varepsilon$ :

$$\lambda_i = \beta \times \left( (1 - \varepsilon) i_j + \varepsilon \sum_{k \neq j} i_k \right) \tag{22.21}$$

The  $(1 - \varepsilon)$  term ensures that the basic reproduction  $R_0$  stays constant. Other spatial models are not based on the theory of metapopulations and instead consider the spatial dimension as a continuous variable. Those models are based on the reaction–diffusion equations that, for simplicity, we will not treat in the present chapter.

### 22.7.4 Stochastic Endemic Models

Sections 22.5 and 22.6 have primarily focused on deterministic models, that is, models in which nothing is random. These models produce pretty good predictions as long as the popu-



**Fig. 22.8.** In a metapopulation context the matrix of contact can be modeled as inversely proportional to the distance between the communities represented by circles.

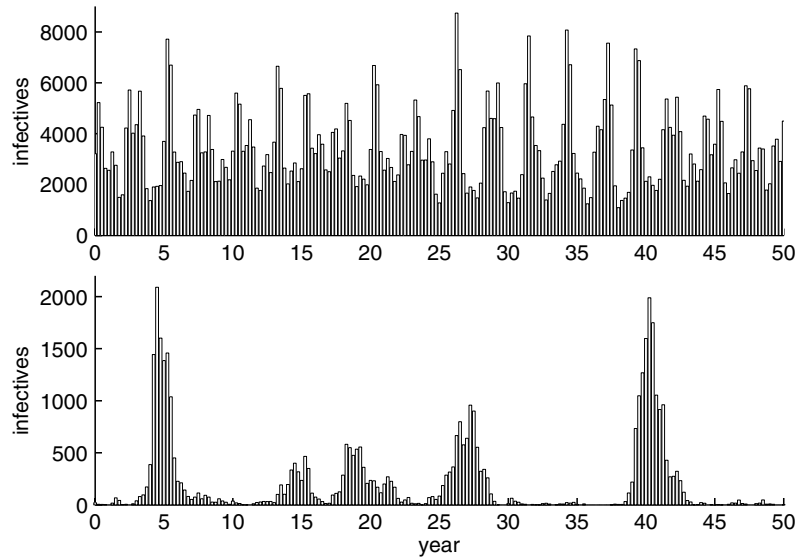
lation is large enough for the stochasticity to have little influence. However, in Sections 22.5.5 and 22.6.3 we highlighted that in small populations, deterministic model predictions become unreliable. To study disease dynamics in small populations, one thus often need a stochastic instead of a deterministic model [6]. In this section we present an easy way to construct a stochastic version of the *SIR* endemic model and we will show that stochasticity introduces enough heterogeneity in the model to produce sustained oscillations.

A stochastic version of the endemic *SIR* model passes through the definition of a Markov process, that is, a process in which the future is independent of the past, given the present. The state space of this process is defined by the number of individuals in each of the three classes susceptible (*S*), infectious (*I*), recovered (*R*). Changes in the state space are characterized by transition events which are listed in Table 22.1. Each transition event occurs with a probabilistic rate derived from the rates of the deterministic model. For example, the probabilistic rate corresponding to an event of birth is defined as follows:

$$P\{1 \text{ birth in } (t, t + \Delta t) \mid S(t) = n\} = \mu n \Delta t + o(\Delta t) \tag{22.22}$$

$$\text{with } \lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0.$$

For numerical simulation, the basic procedure consists in, first, searching the time of the next event (whatever its nature) and, second, determine the nature of this event. As all events are independent, the probabilistic rate that an event occurs, whatever its nature, is simply equal to the sum of the probabilistic rates of all the possible events  $r = \sum_i r_i$ . As future events are independent on past events, the time to the next event follows a negative exponential distribution of parameter  $r$  (see Box 22.3). Thus, the time to the next event can simply be determined by a random realization of a negative exponential probability distribution of parameter  $r$ . Then, the nature of this event is simply determined by a random realization of a multinomial probability distribution of parameters  $r_1/r, r_2/r$ , and so on. This process is reiterated for the duration desired. Figure 22.9 shows results of numerical simulations. Note the resemblance with real time series (compare with subplots of Figures 22.6 and 22.10).



**Fig. 22.9.** Stochastic realizations of an SIR model in populations of 1,000,000 (top) and 100,000 (bottom) individuals. Each bar represents the incidence for one trimester.

## 22.8 DATA ANALYSIS

So far we have presented a variety of disease models, each with its advantages and disadvantages. It should be clear now that there is no one model which is, in absolute, better than the others. The best model depends on the question under investigation. This section is more oriented toward the epidemiological data. We will first see how model parameters can be estimated from the data and then focus specifically on the analysis of longitudinal data by presenting the basic tools of time series analysis.

### 22.8.1 Parameter Estimations

We have seen that mathematical models are characterized by a certain combination of parameters, each with a biological significance such as the force of infection, the birth rate, and so on. We have also seen that mathematical models allow the derivation of concepts which are not directly perceptible on the data, such as the basic reproduction number. In this section we are interested in trying to evaluate the numerical values of these quantities.

All the model parameters can be estimated by maximizing a model likelihood on real data. This procedure is largely used in modeling – not only in epidemiology – and its basic principles are presented in Box 22.2. Parameter estimation by maximization of the likelihood takes into account an error structure and thus allows giving a confidence interval on the estimation. This is one major advantage of the likelihood methods.

We will not present again the likelihood method here (see Box 22.2 for more details). Instead, we are interested in this section on the derivation of parameter values, almost from direct reading from the data, after playing a little bit with the model equations. Contrary to the likelihood methods, this

method does not produce a confidence interval on the parameter estimation – though such an interval can be produced by Monte Carlo simulations (see Section 22.8.1.5). However, this method of parameter estimation is easy, direct, and much faster to implement than the likelihood methods.

**22.8.1.1 The basic reproduction ratio  $R_0$**  In the previous sections we have seen two expressions of the basic reproduction number which can all help to estimate it from the data. The first one evidenced in Section 22.6.1 is relative to the endemic equilibrium value of the proportion of susceptibles in the population. Indeed, by searching the equilibrium point of the system of differential equations we ended with the fact that, at endemic equilibrium, the proportion of susceptibles in the population is equal to the inverse of the basic reproduction number. It is intuitively expected that the higher the basic reproduction number, the lower the proportion of susceptibles at endemic equilibrium in the population. A standard serological survey can easily determine the proportion  $s^*$  of susceptibles of an endemic disease. From this proportion one can thus determine the basic reproductive ratio simply as

$$R_0 = \frac{1}{s^*} \quad (22.23)$$

Such estimations of the basic reproduction number of a variety of viral and bacterial infections are listed in Table 22.2. In Section 22.6.2 we showed an even simpler form of the basic reproduction ratio, provided we have the age of each case notification. From these data one can easily calculate the mean age at infection  $A$ . The life expectancy  $L$  is a demographic information that is available for many human populations. Dividing it by the mean age at infection produces a

**TABLE 22.2.** Some Disease Parameter Values Taken from the Literature. All Parameters are Estimated in Western Countries Unless Otherwise Specified

Diseases	$\gamma^a$	$A^b$	$R_0^c$	$p_c^d$	$T_{obs}^e$	$T_{calc}^f$
Measles	6–7	4–6 <sup>g</sup> , 1–3 <sup>h</sup>	16–17	90–95%	1–2	1–2
Mumps	4–8	6–7	7–8 <sup>g</sup> , 11–14 <sup>h</sup>	85–90%	3, 2–4	3, 2–4
Whooping cough	7–10	4–5	16–17	90–95%	3–4	3–4
Rubella	11–12	9–10 <sup>g</sup> , 2–3 <sup>h</sup>	6–7 <sup>g</sup> , 15–16 <sup>h</sup>	82–87%	3.5	4–5
Chickenpox	10–11	6–8	7–8 <sup>g</sup> , 10–12 <sup>h</sup>	85–90%	2–4	3–4
Smallpox	—	—	—	70–80%	5	4–5
Malaria	—	—	—	99%	—	—

<sup>a</sup>Recovery rate (data from [10, 17, 25]).

<sup>b</sup>Mean age at infection (data form [5]).

<sup>c</sup>Basic reproduction ratio (data from [3, 5, 49]).

<sup>d</sup>Critical mass vaccination coverage [7].

<sup>e</sup>Observed interepidemic period (data from [5]).

<sup>f</sup>Model-predicted interepidemic period (data from [5]).

<sup>g</sup>Western countries.

<sup>h</sup>Developing countries.

good estimate of the basic reproduction ratio of a disease in a given population:

$$R_0 = \frac{L}{A} \tag{22.24}$$

Again this relation seems reasonable as it is intuitively expected that the higher the basic reproduction number  $R_0$ , the lower the mean age at infection  $A$ .

**22.8.1.2 The force of infection  $\lambda$**  In Section 22.6.2 too we arrived at the intuitively plausible conclusion that the mean age at infection is the reciprocal of the force of infection. Thus, knowing the age of the disease cases, one can easily calculate the mean age at infection  $A$  and deduce the force of infection:

$$\lambda = \frac{1}{A} \tag{22.25}$$

For the cases dealing with an age-structured model as in Section 22.7.2, we need to evaluate the force of infection by age cohort. By definition, the force of infection is the probability for a susceptible to get infected. As the events of disease transmission are independent, the number of susceptible follows negative exponential distribution of parameter equal to the force of infection.<sup>2</sup> Said in other words the ratio  $S(a+1)/S(a)$  decreases exponentially at a rate equal to the force of infection:

$$\frac{S(a+1)}{S(a)} = \exp(-\lambda(a)) \tag{22.26}$$

<sup>2</sup>Incidentally, we find again Equation (22.25). Indeed the mean of a negative exponential distribution is, by definition, the inverse of the distribution parameter (see Box 22.3). Thus the mean age at infection is the reciprocal of the force of infection, as in Equation (22.25).

From Equation (22.26) the force of infection by age can thus be easily calculated as long as we are in the possession of disease prevalence  $I(a)$  by age cohorts. Indeed, the number of susceptibles at age 0 is simply equal to the number of newborns ( $S(0) = \mu N$ ) and the other values of  $S(a)$  are then obtained recursively:

$$S(a+1) = S(a) - I(a) \tag{22.27}$$

**22.8.1.3 The coefficient of transmission  $\beta$**  In the case of airborne diseases, the force of infection  $\lambda$  is formally related to the coefficient of transmission  $\beta$ . The simplest of such relations is a linear one (see Section 22.5.1):

$$\lambda = \beta i \tag{22.28}$$

The coefficient of transmission  $\beta$  can thus be estimated from the value of the force of infection  $\lambda$ , as estimated in the above section 22.8.1.2, and the prevalence  $i$ .

For age-structured models, things become a little bit tougher. Indeed, from Equation (22.20) we have a system of  $n$  equations with  $n^2$  unknown variables  $\beta_{i,j}$ :

$$\hat{\lambda}_i = \sum_{j=1}^n \hat{\beta}_{i,j} \hat{i}_j \quad i = 1, \dots, n \tag{22.29}$$

In order to solve this system, it is necessary to formulate hypotheses allowing us to decrease the number of unknown variables down to  $n$ . The first of these hypotheses is an assumption of symmetry [7]:

$$\hat{\beta}_{i,j} \equiv \hat{\beta}_{j,i}, \quad (i,j) \in \{1, \dots, n\}^2 \tag{22.30}$$

However, this hypothesis has the effect of decreasing the number of unknown variables only to  $n(n+1)/2$ . An alternative

consists in defining  $\hat{\beta}_{i,j} \equiv \hat{\beta}_i$  [7]. In any case these hypotheses necessitate the definition of a WAIFW (who acquire the infection from whom) matrix. Considering  $n = 5$  age cohorts, the four most usually used WAIFW matrices are the followings [7]:

$$\begin{aligned} \text{WAIFW}_1 &= \begin{bmatrix} \beta_1 & \beta_1 & \beta_3 & \beta_4 & \beta_5 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 \\ \beta_3 & \beta_3 & \beta_3 & \beta_4 & \beta_5 \\ \beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_5 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \end{bmatrix} \\ \text{WAIFW}_2 &= \begin{bmatrix} \beta_1 & \beta_1 & \beta_1 & \beta_4 & \beta_5 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 \\ \beta_1 & \beta_3 & \beta_3 & \beta_4 & \beta_5 \\ \beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_5 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \end{bmatrix} \\ \text{WAIFW}_3 &= \begin{bmatrix} \beta_1 & \beta_1 & \beta_1 & \beta_1 & \beta_1 \\ \beta_2 & \beta_2 & \beta_2 & \beta_2 & \beta_2 \\ \beta_3 & \beta_3 & \beta_3 & \beta_3 & \beta_3 \\ \beta_4 & \beta_4 & \beta_4 & \beta_4 & \beta_4 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \end{bmatrix} \\ \text{WAIFW}_4 &= \begin{bmatrix} \beta_1 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \\ \beta_5 & \beta_2 & \beta_5 & \beta_5 & \beta_5 \\ \beta_5 & \beta_5 & \beta_3 & \beta_5 & \beta_5 \\ \beta_5 & \beta_5 & \beta_5 & \beta_4 & \beta_5 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \end{bmatrix} \end{aligned}$$

where classes 1, 2, 3, 4, and 5 usually refer to the 0–4 years, 5–9 years, 10–14 years, 15–19 years and 20 years more, respectively. Each of those matrices yields a system of  $n$  equations with  $n$  unknown. The  $\hat{\beta}_i$  can thus be determined from the observed mean incidences by age cohort ( $i$ ) and the mean forces of infection by age cohort calculated in the above Section 22.8.1.2.

For a temporally varying coefficient of transmission, there is no other means of estimation than maximum likelihood, whatever the exact form of the variation on  $\beta$ .

**22.8.1.4 The rate of recovery  $\gamma$**  Quite generally, under steady-state conditions, the quantity in a given compartment is equal to the product of the rate of inflow times the expected sojourn time. In the case of the infective compartment of an SIR model this remark translates into *incidence*  $\times$  *expected sojourn time* = *prevalence* [19]. As mentioned in Section 22.6.1, when the outflow of a compartment occurs at a constant rate, the sojourn time in the compartment follows a negative exponential distribution with parameter equal to the rate of outflow. The mean of a negative exponential distribution is equal to the reciprocal of its parameter, thus the expected sojourn time in the infective compartment is equal to the inverse of the recovery rate. In consequence, we end up with the following:

$$\gamma = \frac{\text{incidence}}{\text{prevalence}} \tag{22.31}$$

Epidemiological data generally contain either incidence or prevalence. However, one can be easily calculated from the other as incidence is equal to the variation of prevalence.

**22.8.1.5 Monte Carlo simulations** We have presented here simple methods to estimate the values of both model parameters (such as the rate of recovery) and emerging quantities (such as the basic reproduction number). These estimations are fast and easy to implement with most available epidemiological data. However, and contrary to likelihood methods, they do not provide any confidence interval on the estimation. One classic method to cope with this is to use Monte Carlo simulations.

The idea of Monte Carlo simulations is to generate a distribution of a parameter by resampling the data [45]. A confidence interval can then be found based on this distribution. In practice, the generation of such a distribution is done as follows. (i) an artificial data of the same length as the original data set is generated by sampling with replacement in the original data set. (ii) The parameter is estimated on this artificial data set and its value kept in memory. Steps (i) and (ii) are repeated a large number of times and the values of the parameters estimated on each artificial data set give a distribution of the parameter. From this distribution one can find a confidence interval. One crucial point in Monte Carlo simulations is related to the choice of the number of time steps (i) and (ii) should be repeated. This number should be large enough for the generated parameter distribution to be considered in a steady state. One way to check for the convergence of the distribution toward a steady state is to follow the evolution of the value of one distribution's statistic (such as the mean) at each new repetition and stop when this statistic seems to have converged to a steady value.

## 22.8.2 Tools for Time Series Analysis

Longitudinal epidemiological surveys produce time series. The object of time series analysis is to look for periodic patterns in the data. Because of the time component, data in time series are not independent. The consequence is that the classic statistical tools that assume independence of data cannot be used on time series [15]. In this section we briefly present the basic tools of time series analysis, from the simplest to the most recent and elaborated.

**22.8.2.1 Stationary time series** The first two methods require the time series to be in a stationary state. A time series is said to be in a stationary state if there is no systematic change in mean (no trend) and in variance [15]. This basically supposes that the signal has constant period and amplitude, which is not the case for numerous real time series. The trend can be removed by considering the residuals from a regression or a nonparametric smoothing such as a B-spline or Loess regression [15]. Another mean for removing the trend consists in applying a moving average to the series [15]: each point a time  $t$  in the series is replaced by the average of the points between

times  $t - T/2$  and  $t + T/2$ , where the length  $T$  of the moving window is to be defined. When  $T$  increases the edge effects increase too and the averaged series will be reduced from  $T$  data points compared to the original series. Lastly, the trends can be removed by considering the variations in the time series [15]: each data point is replaced by the difference with the previous data point. Square transformation of the data generally has the effect of stabilizing the variance and logarithm transformation is usually used to linearize the data, the three methods presented here requiring linear data sets where the effect is proportional to the cause.

**22.8.2.2 Autocorrelograms** This method applies to stationary time series. The idea is to calculate the correlation between a time series and a lagged copy of itself [15] (Figure 22.10). The autocorrelogram plots the value of an autocorrelation coefficient  $r$  against the value of the lag. At lag = 0 the autocorrelation is by definition equal to 1. When the lag increases the value of the autocorrelation coefficient decreases, then becomes negative, and oscillates around the  $r = 0$  horizontal line. The period of the oscillations of the periodogram is the same as the period of the original signal. Moreover, the oscillations are damped. Indeed, because of the dependency between the points of the time series, the noise accumulate additionally, thus hiding any autocorrelation signal when the lag becomes too large. In the same way one can trace correlograms between two different series (called cross-correlograms) to get an idea of the synchronicity or phase difference between two data sets.

**22.8.2.3 Fourier spectra** Like the autocorrelogram this method applies to stationary time series. The Fourier theorem states that any periodic signal  $s(t)$  of frequency  $F_0$  can be decomposed into a sum of sinusoids [13, 15] (Fig. 22.11):

$$s(t) = a_0 + \sum_{n=1}^{+\infty} [a_n \cos(2\pi F_0 n t) + b_n \sin(2\pi F_0 n t)] \quad (22.32)$$

where  $a_0$  is the mean of the signal and  $a_n$  and  $b_n$  ( $n \in \mathbb{N}$ ) are the Fourier coefficients, basically referring to the weight that each harmonic of frequency  $nF_0$  has in the whole signal. This can also be thought of in terms of the magnitude of the correlation between the signal  $s(t)$  and the sinusoid of frequency  $nF_0$ . The use of complex numbers renders this formula much simpler<sup>3</sup>:

$$s(t) = \sum_{n=-\infty}^{+\infty} c_n e^{jn(2\pi F_0)t} \quad (22.33)$$

where  $j$  is the imaginary number and the Fourier coefficients  $c_n$  are now complex. The advantage of this form is that we have only one coefficient  $c_n$  for each frequency  $nF_0$ . Decomposing a time series into a Fourier sum consists in

estimating the coefficients of the Fourier sum. For Equation (22.33) the coefficients are equal to

$$c_n = \frac{1}{T_0} \int_0^{T_0} s(t) e^{-j2\pi F_0 n t} dt \quad (22.34)$$

It is important to realize that the time series  $s(t)$  and the series of Fourier coefficients  $c_n$  describe exactly the same reality, the time series in the time domain and the Fourier coefficients in the frequency domain. The Fourier transform of Equation (22.34) allows passing from the time to the frequency domains whereas the reverse Fourier transform of Equation (22.33) does the opposite transformation. The time or the frequency domains are thus two different ways of looking at the same reality. As in time series analysis we are interested into the regular patterns of a series, it is often more convenient to work into the frequency domain (at least when the series are stationary). This decomposition of a periodic signal into a sum of sinusoids can be generalized to the decomposition of an aperiodic signal where an aperiodic signal is simply a periodic signal with a period equal to  $\infty$  [13,15]. Equation (22.34) then takes the more general form

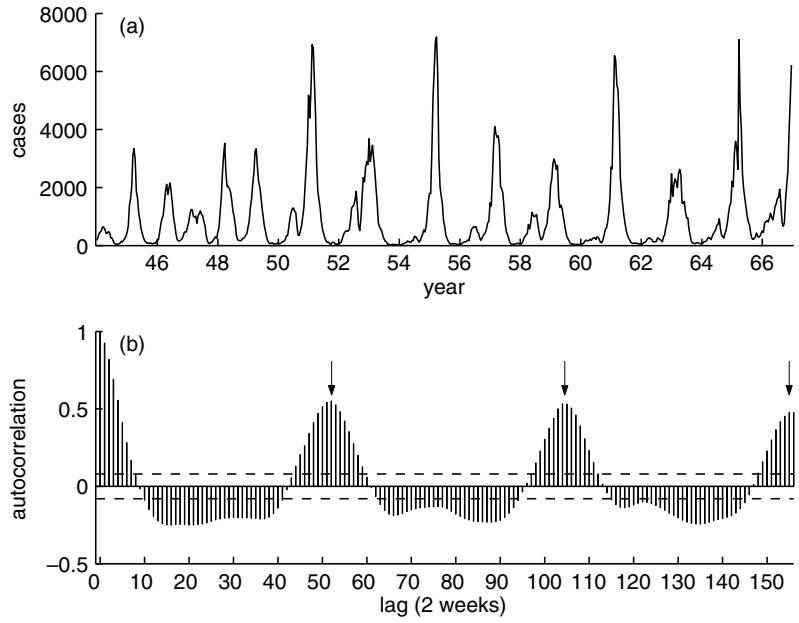
$$S(f) = \int_{-\infty}^{+\infty} s(t) e^{-j2\pi f t} dt \quad (22.35)$$

where the Fourier transform  $S(f)$  is now a continuous function of frequency  $f$ . A Fourier spectrum plots the values of  $S(f)$  against  $f$ . Inspection of such a spectrum gives a clear idea of which frequencies contribute the most to the signal.

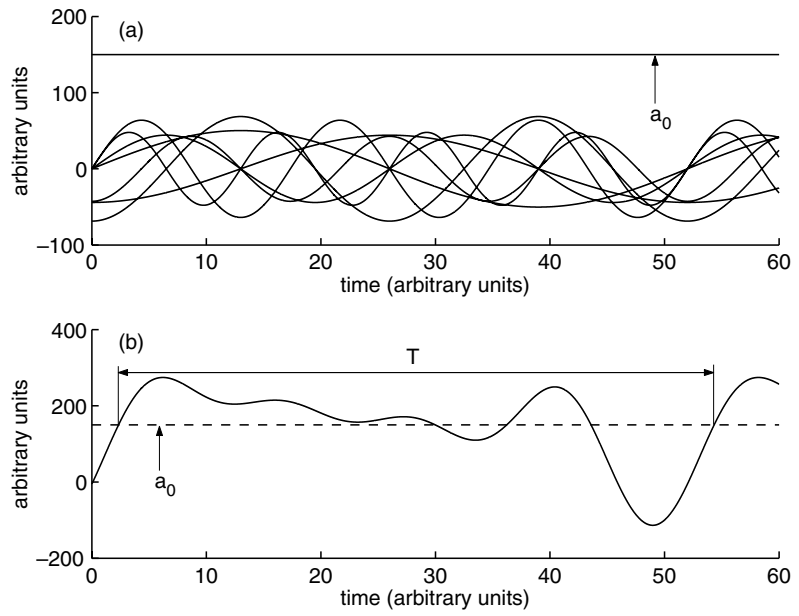
**22.8.2.4 Wavelet analysis** Direct and inverse Fourier transforms force us to visualize a stationary time series either in the time or the frequency domain. Analyzing nonstationary time series, one may, however, be interested in visualizing it in the time and frequency domain at the same time in order to be able to say that the period of the signal is equal to  $T_1$  between times  $t_1$  and  $t_2$ ;  $T_2$  between times  $t_2$  and  $t_3$ ; and so on. One first attempt into this direction has been the use of the Fourier transform on a moving window. The major disadvantage of this *ad hoc* method is that the fixed size of the windows gives different weights to the different frequencies. This inconvenience has been coped by the invention of the wavelets.

The last decade has witnessed the emergence of an impressive number of wavelets. Certainly, the most used in ecology is the Morlet one which is essentially a complex exponential with a Gaussian envelope. The key advantage of wavelets relative to sinusoids used in Fourier analysis is that not only they can be moved along the signal (as in windowed Fourier analysis) but also they can be stretched to account equally for the different frequencies [13,58]. A wavelet spectrum is thus a three-dimensional graph which plots the correlation of the wavelet with the signal as a function of both the location of the wavelet along the signal (time domain) and the stretching of the wavelet (frequency domain). Figure 22.12 shows an example for the measles

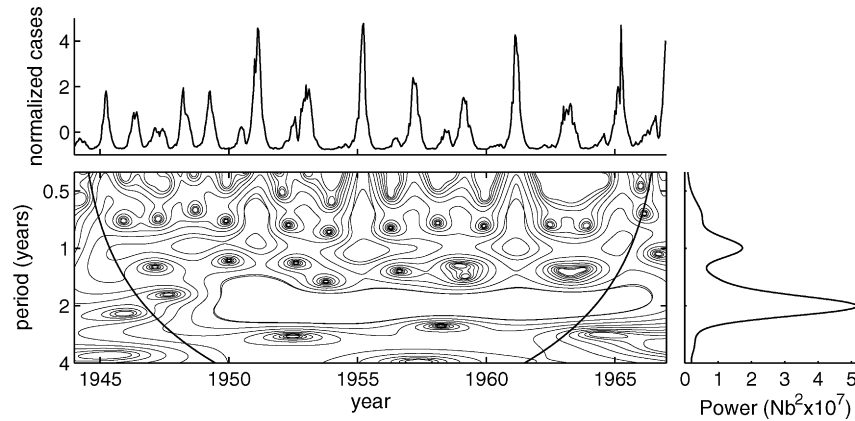
<sup>3</sup>After the application of the key mathematical relationship:  $e^{j\theta} = \cos \theta + j \sin \theta$ .



**Fig. 22.10.** Autocorrelation plot of the biweekly measles notification cases for London from 1944 to 1966. (a) Time series of the cases. The correlation between the time series and a lagged copy of itself is calculated. As the notifications are biweekly, the lag is equal to 2 weeks. (b) The autocorrelogram shows the value of the autocorrelation against the lag. The dashed line represent the 95% confidence limits about zero. When lag = 0 the series is correlated with itself and thus the autocorrelation is equal to 1. The autocorrelation coefficient then reaches maxima at lag = 53, 105, and 156 (106, 210, 312 weeks, respectively, see vertical arrows), thus evidencing a period of about 2 years. Data downloaded from <http://www.zoo.cam.ac.uk/zoostaff/grenfell/measles.htm> [27].



**Fig. 22.11.** Decomposition of a periodic signal into Fourier sum of sinusoids. (a) Plot of the sinusoidal component of the whole signal (b). With reference to Equation (22.32)  $a_0 = 150$ ,  $a_1 = 244.16$ ,  $b_1 = 20.22$ ,  $a_2 = 268.66$ ,  $b_2 = 44.22$ ,  $a_3 = 242.34$ ,  $b_3 = 63.82$ ,  $a_4 = 0$ ,  $b_4 = 47.83$ ,  $T = 1/F_0 = 52$  arbitrary units.



**Fig. 22.12.** Wavelet power spectrum of the measles notification cases for London between 1944 and 1966. The magnitude of the correlation between the series and the wavelet increases from blue to black. The parabolic curves represent the cone of influence. Because of edge effects, everything below the cone of influence cannot be interpreted. The above graph shows the time series of normalized measles cases and the right graph is the Fourier spectrum. It is clear that the wavelet power spectrum combines information of both time and frequency domains. Data downloaded from <http://www.zoo.cam.ac.uk/zoostaff/grenfell/measles.htm> [27]. See color plates.

cases of London from 1944 to 1966. With such a graph one is able to say the frequency of a signal at any time.

## 22.9 APPLICATIONS TO VACCINATION POLICIES

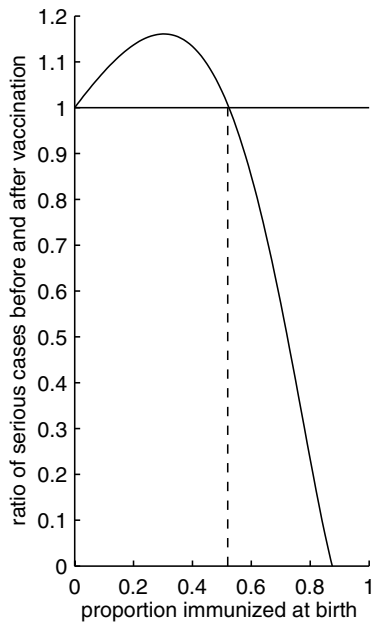
After an overview of the basic epidemiological models and results, as well as statistical tools for the epidemiologist, the last section is devoted to practical applications for the development of vaccination policies. In public health, vaccination policies are decisions made by governments and applied on large spatial and temporal scales. The ultimate aim of a vaccination policy is the eradication of a disease from a population. This goal is extremely difficult to achieve in practice and most vaccination strategies are imperfect in the sense that they only decrease (sometimes dramatically) the number of cases, without, however, eradicating the disease [7]. In this context, vaccination can yield side effects on the disease statics and dynamics that are important to evaluate. There currently exist two major vaccination strategies – the mass vaccination, which is the most ancient and still the most applied, and the recently developed pulse vaccination which is used in an increasing number of countries.

### 22.9.1 Mass Vaccination Strategy

Mass vaccination strategy is the most ancient and still the most widely used vaccination scheme. It consists in vaccinating a large proportion of infants before the mean age at infection [7], for example, the 0–2 age cohort for the measles-mumps-rubella (MMR) vaccine in the United States. Its first applications started in the sixties against measles in the North American and European countries where they have caused a dramatic decrease of the number of cases.

**22.9.1.1 Calculating the vaccination coverage** The derivation of the optimal vaccination coverage is based on the properties of the endemic equilibrium. At equilibrium the replacement number  $R$  (see Section 22.5.3) is equal to the basic reproduction number  $R_0$  times the proportion of susceptibles:  $R = R_0 s^*$ . Applying a vaccination coverage equal to  $p$  has the effect of diminishing the proportion of susceptibles by  $p$ :  $s^* = 1 - p$ . A condition for disease eradication is that the reproduction number be less than 1:  $R = R_0 s^* = R_0(1 - p) < 1$ , or  $p > 1 - 1/R_0$ . Thus the critical vaccination coverage  $p_c$  for disease eradication is  $p_c = 1 - 1/R_0$  [7]. Note that this result shows that we do not need to vaccinate each individual to protect the whole population. Note too that this property known as herd immunity is not evident from the data and emerges only from the model. The higher the basic reproduction number, the higher the vaccination coverage should be. Vaccination coverages of major human infectious diseases are given in Table 22.2. We can see that many infectious diseases require vaccination coverage which are far too higher to be achieved in practice. This is further complicated by other mechanisms such as the vaccine efficacy. Consider, for example, measles and rubella for which estimates of the critical vaccination coverage based on  $R_0$  are 0.94 and 0.86, respectively. A vaccine efficacy of 0.95 means that 5% of those vaccinated do not become immune. In consequence, taking into account vaccine efficacy necessitates coverages of 0.99 and 0.91 for measles and rubella, respectively [34]. This explains why the only human infectious disease which has been eradicated successfully worldwide so far is smallpox which has the lowest critical vaccination coverage.

**22.9.1.2 Consequences on the statics** A first and expected effect of vaccination is that fewer people will



**Fig. 22.13.** Model predicted effect of a mass vaccination policy at birth on the number of congenital rubella syndromes (CRS). The graph plots the ratio of CRS before and after the start of the mass vaccination policy against the vaccination coverage. It shows that low vaccination coverages (less than 50% here) should be avoided as they can increase the absolute number of serious cases (model from [4]).

experience infection. But the decrease of the force of infection due to vaccination means that the mean age at infection of the smaller number of people who do acquire infection increases [7] (recall Equation (22.25)). If the probability of disease complications increases with age, it is thus possible that some vaccination programs could actually increase the absolute number of serious cases. The likelihood of such a perverse outcome again can only be evaluated thanks to disease models. A classical example is the one of the congenital rubella syndrome (CRS), treated in detail by Anderson and May [4]. They have evidenced that the absolute number of CRS can actually increase with the vaccination coverage when the vaccination coverage is low (Figure 22.13).

**22.9.1.3 Consequences on the spatial and/or temporal dynamics** The spatio-temporal dynamics of a disease is of primary interest if we are to design efficient country-wide vaccination policies [28]. For example, the global eradication of a disease would be easier if the local dynamics are synchronous. In the case where local dynamics are completely asynchronous, local extinctions would be quickly followed by migration of infectious individuals from neighbor communities experiencing an epidemic outbreak. With a very simple endemo-epidemic model Earn et al. [22] have shown that the disease dynamics complexity – as given by the length of the period and the number of different attractors (see Box 22.4) – increases with vaccination coverage.

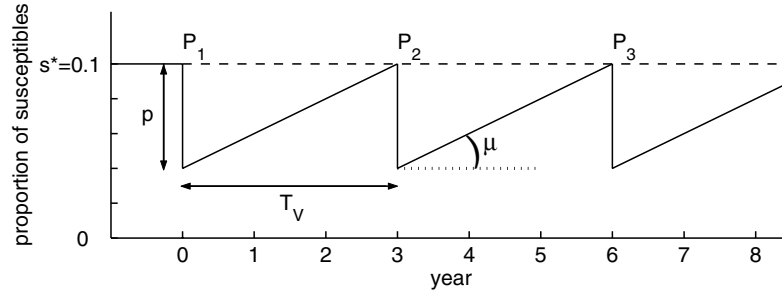
Furthermore, a simple island model such as the one presented in Section 22.7.3 evidences that an increase in the vaccination coverage results in a decrease of the spatial synchrony of disease dynamics. This is intuitive as the coupling between the different subpopulations is assured by the migration of infective (see Equation (22.21)). Vaccination has the effect of decreasing the number of infectives and thus the synchrony between the different subpopulations. In conclusion, mass vaccination has the effect of (i) increasing disease dynamics complexity and (ii) desynchronizing local dynamics. The first consequence accentuates the second as the probability of dynamics synchronicity naturally decreases with the level of complexity. These model predictions have been successfully confirmed on real data analysis [52]. As the mass vaccination tends to desynchronize the spatial dynamics of the disease, it renders global disease eradication even more difficult to achieve in practice than expected from the above theoretical predictions (Section 22.9.1.1).

## 22.9.2 Pulse Vaccination Strategy

As highlighted in the previous Section 22.9.1, mass vaccination strategy requires a too high systematic vaccination coverage to be achieved in practice. A recently proposed and potentially less expensive strategy is vaccination in pulses [1, 50]. This approach consists in vaccinating a certain proportion of the population at regular intervals of time. The rationale behind this is to vaccinate sufficiently and frequently enough to maintain the percentage of susceptibles below the threshold necessary for an epidemic to start (Figure 22.14). What makes this policy less expensive than the mass vaccination strategy is that it explicitly accounts for the dynamics of the host population through the birth rate. Several theoretical works have been carried out to express the optimal vaccination coverage and frequency as a function of the host demographic characteristics [21,55,57]. The simplest one is derived from the Pythagore theorem (see Figure 22.14). This theory has been successfully applied in campaigns against poliomyelitis and measles in Central and South America, and measles in the United Kingdom in 1994.

**22.9.2.1 Spatial dynamics** Using a simple endemo-epidemic model Earn et al. [23] have shown that a same pulse vaccination strategy tends to synchronize disease dynamics in independent localities (see Figure 22.15). This phenomenon is a case of Moran effect where the same causes produce the same effects [12, 30, 47]. Indeed, by its periodical nature, an imperfect pulse vaccination strategy acts as a forcing driving the disease dynamics. Two independent populations submitted to the same pulse vaccination scheme will thus exhibit similar and synchronous disease dynamics. The effect of pulse vaccination on the spatial dynamics of a disease is thus opposite to the one of mass vaccination and this facilitates the achievement of a global disease eradication [38].

**22.9.2.2 Resonance** A second side-effect that can be associated to the periodic nature of pulse vaccination is the phenomenon of resonance. Theory of oscillator dynamics predicts



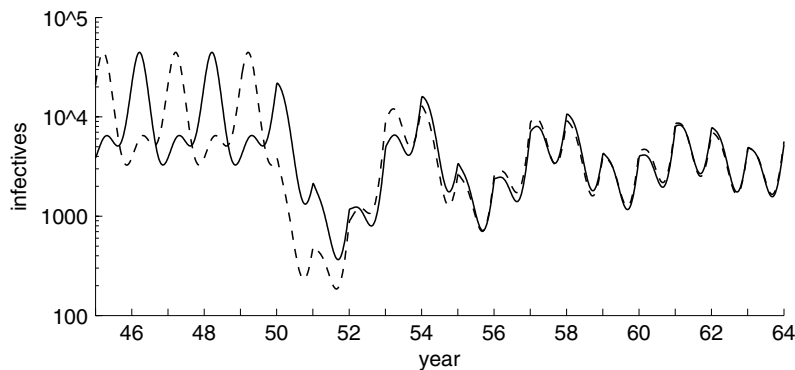
**Fig. 22.14.** Pulse vaccination scheme. The graph shows the proportion of susceptibles as a function of time. The aim is to find a vaccination coverage  $p$  and a frequency  $1/T_V$  of the vaccination pulses  $P_1$ ,  $P_2$ ,  $P_3$ , and so on, such as the proportion of susceptibles stays below the critical value  $s^*$  necessary for an epidemic to start (represented by the dashed line). Births fill the stock of susceptibles at a constant rate  $\mu$ .

that an oscillating dynamical system (such as an epidemiological one) submitted to a periodic forcing (such as an imperfect pulse vaccination) can produce phenomena of resonance [36]. Resonance is a generic term indicating that the amplitude of observed oscillations depends on the period of the forcing, and has a maximum, called peak of resonance. These theoretical predictions and their epidemiological consequences have been investigated on a disease system by numerical simulations and data analyses [16]. Figure 22.16 shows that the mean annual number of infectives globally decreases as the frequency of vaccination pulses increases. However, resonance is responsible for the peaks observed on this general trend. The major one occurs at a vaccination frequency close to 2 years, the others simply being harmonics of it. The practical consequence of these peaks is that, locally on the vaccination frequency dimension, the mean annual number of infectives counter-intuitively increases with the frequency of vaccination.

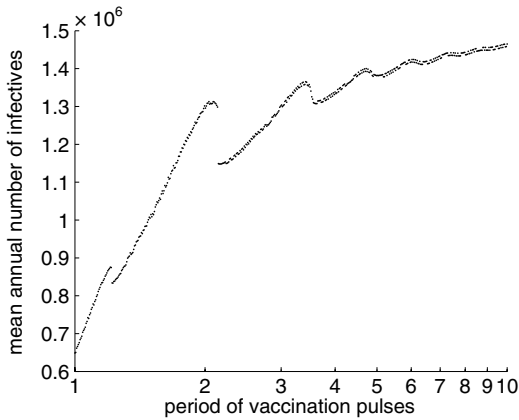
## 22.10 CONCLUSION

### 22.10.1 What We Have Seen

Statistical analyses of epidemiological data help to characterize, quantify, and summarize the way diseases spread in host populations (Section 22.8). The aim of epidemiological modeling is to understand the behavior of diseases in nature (Section 22.2). Because of ethical and practical impossibility to perform experiments in public health, mathematical models appear as a cheap and efficient way to explore and test hypotheses (Section 22.2.2). In addition to force the investigator to think rigorously (Sections 22.5.1 and 22.5.2, Box 22.3), models provide powerful conceptual results such as the basic reproduction number and threshold effects (Section 22.5.3), or the herd immunity (Section 22.9.1.1). Even if very interesting pure theoretical works have been realized, the key element of epidemiological modeling is to



**Fig. 22.15.** Effect of a same pulse vaccination strategy on independent disease dynamics. The simulations start so that the disease dynamics are in opposition of phase. At year 50 a pulse vaccination strategy is started with  $p = 20\%$  and  $T_V = 1$  year, progressively synchronizing the two dynamics.



**Fig. 22.16.** Mean annual number of infectives as a function of the period  $T_V$  of vaccination pulses. The general trend is a decrease of the mean annual number of infectives as the frequency of vaccination ( $1/T_V$ ) increases. However, resonance is responsible for these peaks on this general trend. The consequence of such peaks is that, locally on the vaccination frequency dimension, the mean annual number of infectives counter-intuitively increases with the frequency of vaccination.

link model with data. Likelihood methods are modern and efficient ways to do so (Box 22.2). Models thus allow to estimate epidemiological parameters and also to identify crucial data that need to be collected.

This chapter was centered on the *SIR* model. Although one of the simplest epidemiological models, it is still one of the most used, particularly to study childhood viral and bacterial infections. There is a multitude of ways to complexify this simple model in order to account for more and more phenomena. However, the more complex is not necessarily the best and it is the purpose of a model that dictates its degree of complexity (Section 22.2, Box 22.1). We have explored models to study single epidemics (Section 22.5), endemic diseases (Section 22.6), spatial disease dynamics (Section 22.7.3), and have illustrated how these mathematical tools can be used for the development of public health policies in helping defining optimal vaccination strategies (Section 22.9).

### 22.10.2 What We Have Not Seen

Of course, the list of what we have not seen about epidemiological models is much longer than the list of what we have glanced at.

Some classes can be added to the simple *SIR* model. For example, a commonly used model for childhood diseases is the *SEIR* one which adds a class of exposed (*E*) individuals, that is, individuals which are infected but not infectious yet. For most childhood infectious diseases the latence phase is often as long as the infectious and should be accounted for as it can substantially change the epidemiological conclusions of the models. *MSEIR* models further add a class accounting for the post-birth period during which newborns are protected from infections by maternal antibodies.

All the models covered in this chapter assume that the host population is of constant size. This, thus, excludes both diseases in exponentially growing populations like in most developing countries and disease-induced mortality as the case for many infections including childhood diseases in developing countries, malaria, and so on. Accounting for a nonconstant host population size requires the explicit modeling of the host population dynamics, in addition to the disease dynamics.

For sexually transmitted diseases (STD), contagious contacts are not established randomly as usually assumed for airborne infections. Besides a strong heterogeneity in the sexual activity, sexual contact occurs preferentially between people of the same sexual activity, creating this core effect in the epidemiology. Models for STD should thus account for all these degrees of heterogeneity in the contacts. In addition, many STD result in little or no acquired immunity following recovery, and *SIRS* models would be more appropriate.

Other particular epidemiological systems requiring adapted models include, among others, mother-to-child diseases for which not all children are born into the susceptible compartment and diseases propagated by syringe-sharing intravenous drug users such as HIV/AIDS.

The epidemiology of multiple-host diseases is far more complicated than the one of directly transmitted infections. Compartmental models of such diseases need to account for the dynamics of the disease in the different hosts or reservoirs, and possibly also for the dynamics of these different hosts or reservoirs. For these diseases, the modeling of passage from one host to the other is not always an easy task.

Lastly, so far we have dealt primarily we microparasitic infections. Contrary to microparasites (Section 22.4), macroparasites refer to large-size parasites (helminths, arthropods) with direct reproduction in the definitive host. Macroparasites generally have longer generation time (often an appreciable proportion of the host life span) than microparasites. The immunity following the recovery from a macroparasitic infection is generally of short duration and the number of parasites per host is a strong determinant of the epidemiology. From a modeling point of view, this means that the simple compartmental models presented in this chapter for microparasites should be replaced by more complicated models accounting for the distribution of parasites among the hosts. This is totally another subject and we refer the reader to the classical book by Anderson and May [7] for more details.

### 22.11 SUMMARY

By clarifying rigorously the assumptions, the variables, and the parameters, mathematical modeling allows understanding the observed spread of diseases in space and time. Epidemiological model further provides important conceptual results including the basic reproduction number, the threshold effects, and the herd immunity. For evident ethical and practical reasons, experiments in public health are often impossible to perform and mathematical models thus appear as a cheap and efficient

way to explore and test hypotheses. This is, for example, of particular practical utility in the design of vaccination policies. One key aspect of epidemiological models is their link to real data. Such data often stand under the form of time series which necessitate specific statistical tools for their analysis. Models can always be complicated to improve their fit to real data. However, more complex models are not always the best and it is the question under investigation that should dictate the optimal level of complexity.

## ACKNOWLEDGMENTS

Benjamin Roche drew Figure 22.9. The authors thank H el ene Broutin, Guillaume Constantin de Magny, Christian Paroissin, and Benjamin Roche for their useful comments on the first versions of the chapter. M.C. is funded by a Lavoisier fellowship from the *French Ministry of Foreign Affairs*, J.F.G. is sponsored by the *Institut de Recherche pour le D veloppement* and the *Centre National de la Recherche Scientifique*. P.R. is funded by the *National Institutes of Health*, the *National Science Foundation*, and the *Ellison Medical Foundation*.

## REFERENCES

- Agur Z, Cojocaru L, Anderson RM, Danon YL. Pulse mass measles vaccination across age cohorts. *Proc Natl Acad Sci USA* 1993;**90**:11698–702.
- Akaike H. Information theory as an extension of the maximum likelihood principle. In: Ptroc BN, Csaki F, eds. *Second International Symposium on Information Theory*. Akademiai Kado, 1973, pp. 267–81.
- Anderson RM. Transmission dynamics and control of infectious disease agents. In: Anderson RM, May RM, eds. *Population Biology of Infectious Diseases*. Springer-Verlag, Berlin, 1982, pp. 149–76.
- Anderson RM, May RM. Vaccination against rubella and measles: quantitative investigations of different policies. *J Hyg Camb* 1983;**90**:259–325.
- Anderson RM, May RM. Vaccination and herd immunity to infectious diseases. *Nature* 1985;**318**:323–9.
- Anderson RM, May RM. The invasion, persistence and spread of infectious diseases within animal and plant communities. *Philos Trans R Soc Lond* 1986;**B314**:533–70.
- Anderson RM, May RM. *Infectious Diseases of Humans. Dynamics and Control*. Oxford University Press, Oxford, 1991.
- Bailey NJT. *The Mathematical Theory of Infectious Diseases and its Application*. Griffin, London, 1957.
- Bartlett MS. The critical community size for measles in the United States. *J R Stat Soc* 1960;**123**:37–44.
- Benenson AS. *Control of Communicable Diseases in Man*. American Public Health Association, Washington, D.C., 1975.
- Bernoulli D. Essai d’une nouvelle analyse de la mortalit e caus ee par la petite v erole et des avantages de l’inoculum pour la pr evenir. *M m Math Phys Acad Roy Sci Paris* 1760;1–45.
- Blasius B, Huppert A, Stone L. Complex dynamics and phase synchronization in spatially extended ecological systems. *Nature* 1999;**399**:354–9.
- Burke Hubbard B. *The World According to Wavelets: The Story of a Mathematical Technique in the Making*. AK Peters, 1998.
- Burnham KP, Anderson DR. *Model Selection and Multimodel Inference: a Practical Information-theoretic Approach*. Springer-Verlag, Berlin, 2002.
- Chatfield C. *The Analysis of Time Series: an Introduction*. Chapman & Hall, London, 1996.
- Choisy M, Gu egan JF, Rohani P. Dynamics of infectious diseases and pulse vaccination: teasing apart the embedded resonance effects. *Physica D* 2005;**223**:26–35.
- Christie AB. *Infectious Diseases: Epidemiology and Practice*. Churchill Livingstone, London, 1974.
- Daley DJ, Gani J. *Epidemic Modelling: an Introduction*. Cambridge University Press, Cambridge, 1999.
- Diekmann O, Heesterbeek JAP. *Mathematical Epidemiology of Infectious Diseases. Model Building, Analysis and Interpretation*. Wiley and Sons, Chichester, 2000.
- Dietz K. The incidence of infectious diseases under the influence of seasonal fluctuations. *Lect Notes Biomath* 1976;**11**:1–5.
- d’Onofrio A. Stability properties of pulses vaccination strategy in SEIR epidemic model. *Math Biosci* 2002;**179**:57–72.
- Earn DJD, Rohani P, Bolker BM, Grenfell BT. A simple model for complex dynamical transitions in epidemics. *Science* 2000;**287**:667–70.
- Earn DJD, Rohani P, Grenfell BT. Persistence, chaos and synchrony in ecology and epidemiology. *Proc R Soc Lond* 1998;**B265**:7–10.
- Ellner S, Bailey BA, Bobashev GV, Gallant AR, Grenfell BT, Nychka DW. Noise and nonlinearity in measles epidemics: combining mechanistic and statistical approaches to population modeling. *Am Nat* 1998;**151**:425–40.
- Fenner F, White DO. *Medical Virology*. Academic Press, New York, 1970.
- Fine PEM, Clarkson JA. Measles in England and Wales. I. An analysis of factors underlying seasonal patterns. *Int J Epidemiol* 1982;**11**:5–14.
- Grenfell BT, Bjornstad ON, Finkenst adt B. Dynamics of measles epidemics: scaling noise, determinism and predictability with the TSIR model. *Ecol Monogr* 2002;**72**:185–202.
- Grenfell BT, Bjornstad ON, Kappey J. Travelling waves and spatial hierarchies in measles epidemics. *Nature* 2001;**414**: 716–23.
- Grenfell BT, Harwood J. (Meta)population dynamics of infectious diseases. *Trends Ecol Evol* 1997;**148**:317–35.
- Grenfell BT, Wilson K, Finkenst adt BF, et al. Noise and determinism in synchronized sheep dynamics. *Nature* 1998;**394**:674–7.
- Grossman Z. Oscillatory phenomena in a model of infectious diseases. *Theor Popul Biol* 1980;**18**:204–43.
- Hamer WH. Epidemic disease in England. *Lancet* 1906;**i**:733–9.
- Harris TE. *The Theory of Branching Processes*. Springer-Verlag, Berlin, 1963.
- Hethcote HW. The mathematics of infectious diseases. *SIAM Rev* 2000;**42**:599–653.
- Hilborn R, Mangel M. *The Ecological Detective. Confronting Models with Data*. Princeton University Press, Princeton, 1997.
- Jackson EA. *Perspectives of Nonlinear Dynamics: Volume 1*. Cambridge University Press, Cambridge, 1992.
- Johnson JB, Omland KS. Model selection in ecology and evolution. *Trends Ecol Evol* 2004;**19**:101–8.

38. Keeling MJ, Bjørnstad ON, Grenfell BT. Metapopulation dynamics of infectious diseases. In: Hanski IA, Gaggiotti OE, eds. *Ecology, Genetics and Evolution of Metapopulations. Standard Methods for Inventory and Monitoring*. Elsevier, 2004, pp. 415–45.
39. Keeling MJ, Grenfell BT. Understanding the persistence of measles: reconciling theory, simulation and observation. *Proc R Soc Lond* 2002;**B269**:335–43.
40. Keeling MJ, Rohani P, Grenfell BT. Seasonally forced diseases dynamics explored as switching between attractors. *Physica D* 2001;**148**:317–35.
41. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proc R Soc Lond* 1927;**A115**: 700–21.
42. Kot M. *Elements of Mathematical Ecology*. Cambridge University Press, Cambridge, 2001.
43. Lloyd AL. Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proc R Soc Lond* 2001;**B268**:985–93.
44. London WP, Yorke JA. Recurrent outbreaks of measles, chickenpox and mumps. I. Seasonal variation in contact rates. *Am J Epidemiol* 1973;**98**:453–68.
45. Manly B. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Chapman & Hall, London, 1997.
46. McCallum H, Barlow N, Hone J. How should pathogen transmission be modelled? *Trends Ecol Evol* 2001;**16**:295–300.
47. Moran PAP. The statistical analysis of the Canadian lynx cycle. II. Synchronization and meteorology. *Aust J Zool* 1953;**1**:291–8.
48. Morens DM, Folkers GK, Fauci AS. The challenge of emerging and reemerging infectious diseases. *Nature* 2004;**430**: 242–9.
49. Nokes DJ, Anderson RM. The use of mathematical models in the epidemiology study of infectious diseases and in the design of mass vaccination programmes. *Epidemiol Infect* 1988;**101**:1–20.
50. Nokes DJ, Swinton J. Vaccination in pulses: a strategy for global eradication of measles and polio? *Trends Microbiol* 1997;**5**(1):14–9.
51. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipes in C. The Art of Scientific Computing*. Cambridge University Press, Cambridge, 1997.
52. Rohani P, Earn DJD, Grenfell BT. Opposite patterns of synchrony in sympatric diseases metapopulations. *Science* 1999;**286**:968–71.
53. Rohani P, Keeling MJ, Grenfell BT. The interplay between determinism and stochasticity in childhood diseases. *Am Nat* 2002;**159**:569–481.
54. Ross R. *The Prevention of Malaria*. Murray, London, 1911.
55. Shulgin B, Stone L, Agur Z. Pulse vaccination strategy in the SIR epidemic model. *Bull Math Biol* 1998;**60**:1123–48.
56. Soper MA. The interpretation of periodicity in disease prevalence. *J R Stat Soc* 1929;**A92**:34–61.
57. Stone L, Shulgin B, Agur Z. Theoretical examination of the pulse vaccination policy in the SIR epidemic model. *Math Comput Model* 2000;**31**:207–15.
58. Torrence C, Compo GP. A practical guide to wavelet analysis. *Bull Am Meteorol Soc* 1998;**79**:61–78.
59. Turchin P, Hanski I. Contrasting alternative hypotheses about rodent cycles by translating them into parametrized models. *Ecol Lett* 2001;**4**:267–76.
60. Wearing HJ, Rohani P, Keeling MJ. Appropriate models for the management of infectious diseases. *Public Libr Sci – Med* 2005;**2**(7):e174.