Contents

Mathematical Epidemiology of Infectious Diseases

David J. D. Earn

Mathematical Epidemiology of Infectious Diseases

David J. D. Earn

1. Introduction

Mathematical modelling has contributed significantly to our understanding of the epidemiology of infectious diseases [**AM91**], and is beginning to have a substantial impact on research at the immunological and molecular levels [**NM00**]. The earliest mathematical study of an infectious disease was undertaken more than 200 years ago by Daniel Bernoulli, who attempted to influence public health policy concerning smallpox [**Ber1760, BB04**]. Nevertheless, a firm theoretical basis for infectious disease modelling was not established until the 20th century. Some of the first analyses provided fundamental insights about thresholds for epidemics to occur and the incidence patterns that result [**Ros11, KM27**], while more recent work has elucidated effects of demographic stochasticity [**Bar60, Bai75, Kur80, AB00**], heterogeneity in age, genetics, social and spatial structure [**AM91**]. The ultimate goal of most of this work is to contribute to efforts to develop improved strategies for eradication and control of many infectious diseases.

If we hope to build models that can help predict the effects of changes in control strategies, we must first demonstrate that models can successfully predict patterns of epidemics that have been observed in the past. Fortunately, medical and public health administrators in many countries have kept records of reported deaths from (and cases of) infectious diseases for decades and, in some places, centuries. These records, together with immunological and genetic data that have become available more recently, provide an empirical backbone for the mathematical study of infectious disease transmission dynamics and control.

These lectures will introduce you to the subject of mathematical epidemiology from an applied perspective. By "applied" I mean that we will try to make sense of specific examples of data, and we will try to draw inferences that could influence decisions of public health policymakers. We will begin with a few examples of epidemiological data that present intriguing puzzles, and some elementary statistical tools for quantifying the patterns in these data. We will then examine two very different approaches for investigating these patterns — statistical time series modelling and mechanistic mathematical modelling — before considering how we might

Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada L8S 4K1
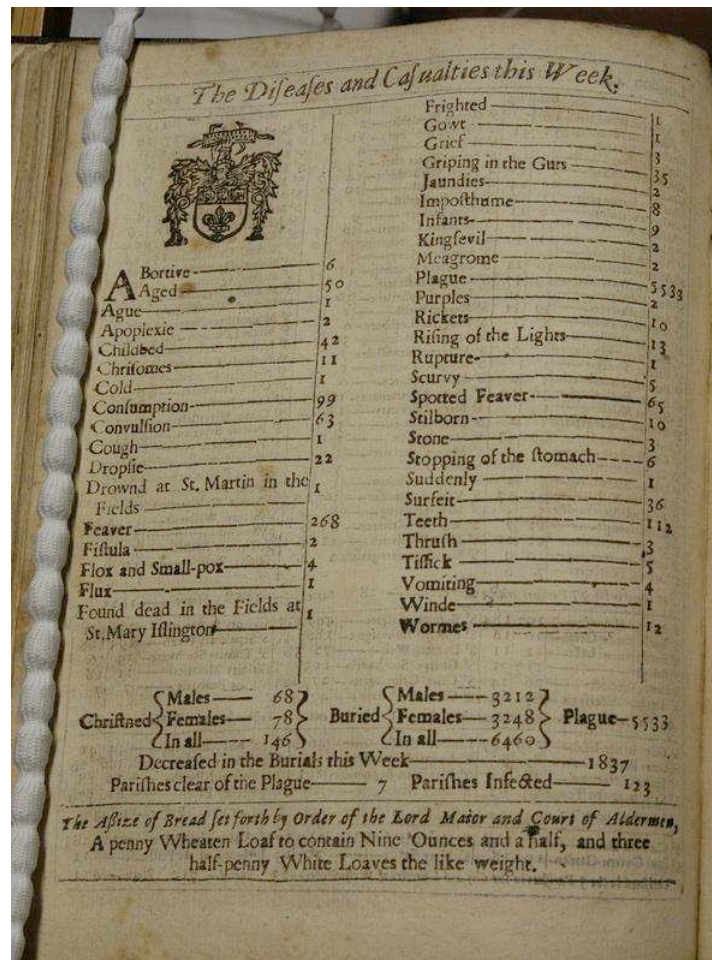
E-mail address: earn@math.mcmaster.ca

predict changes in patterns of epidemics and how we might manipulate infectious disease dynamics so as to reduce expected morbidity and mortality.

2. Describing epidemics

2.1. Plague

One of the most famous examples of an epidemic of an infectious disease in a human population is the Great Plague of London, which took place in 1665 [**MM04**]. We know quite a lot about the progression of the Great Plague because weekly bills of mortality from that time have been retained. A photograph of such a bill is shown in Figure 1. Note that the report indicates that the number of deaths from plague (5533) was more than 37 times the number of births (146) in the week in question, and that wasn't the worst week! An even worse plague — known as the *Black*

FIGURE 1. A bill of mortality for the city of London, England, for the week of 26 September to 3 October 1665. This photograph was taken by Claire Lees at the Guildhall in London, England, with the permission of the librarian.

Death — occurred in the 14th century, but the remaining records of that epidemic are much less detailed.

Putting together the weekly counts of plague deaths from all the relevant mortality bills, we can obtain the *epidemic curve* for the Great Plague, which I've plotted in the top left panel of Figure 2. The bottom left panel of Figure 2 shows weekly mortality from plague in London over a period of 70 years. The Great Plague is the rightmost (and highest) peak in the plot. You can see that on a longer timescale, there was a complex pattern of plague epidemics including extinctions and re-emergences. At present, however, we have no direct evidence that all those plagues were really caused by the same pathogenic organism.
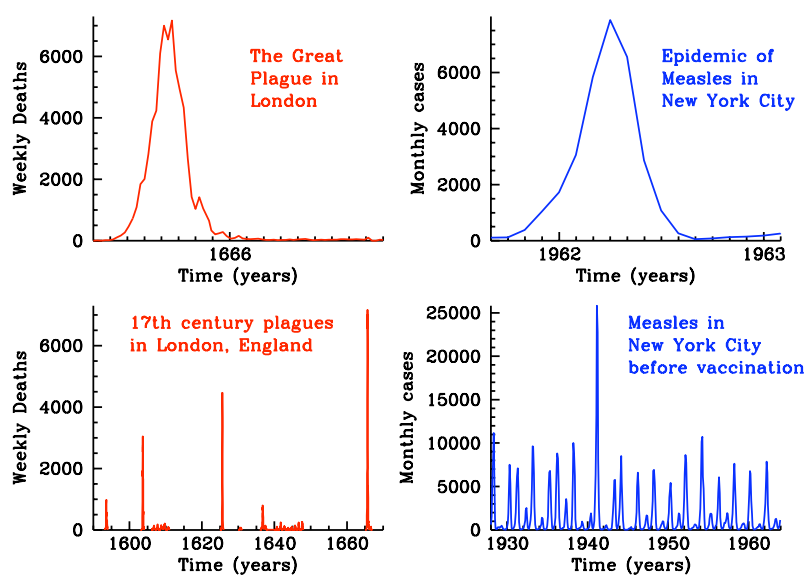
FIGURE 2. Epidemic curves for plague in London (left panels) and measles in New York City (right panels). For plague, the curves show the number of deaths reported each week. For measles, the curves show the number of cases reported each month. In the top panels, the small ticks on the time axis occur at monthly intervals.

2.2. Measles and other childhood diseases

A less contentious example is given by epidemics of measles, which are definitely caused by a well-known virus that infects the respiratory tract in humans and is transmitted by airborne particles [**AM91**]. Measles gives rise to characteristic red spots that are easily identifiable by physicians who have seen many cases, and parents are very likely to take their children to a doctor when such spots are noticed. Consequently, the majority of measles cases in developed countries end up in the office of a doctor (who, in many countries, is required to report observed measles cases to a central body). The result is that the quality of reported measles case data is unusually good.

An epidemic curve for measles in New York City in 1962 is shown in the top right panel of Figure 2. The period shown is 17 months, exactly the same length of time shown for the Great Plague of London in the top left panel. The 1962 measles epidemic in New York took off more slowly and lasted longer than the Great Plague of 1665. The bottom right panel of Figure 2 shows reported measles cases in New York City for a 36 year period, the end of which includes the 1962 epidemic. The cyclicity in the epidemic pattern is striking and much more regular than the series of plague epidemics in the bottom left panel.

There are also striking differences between the epidemic patterns of measles and those of other common childhood diseases. Figure 3 shows examples of sequences of recorded epidemics of measles, whooping cough, rubella and chicken pox. While all of these diseases display recurrent epidemics, the detailed structure of the temporal patterns could hardly be more varied. This is especially puzzling given that all these diseases are caused by respiratory infections with many clinical similarities, and all are observed primarily in children.
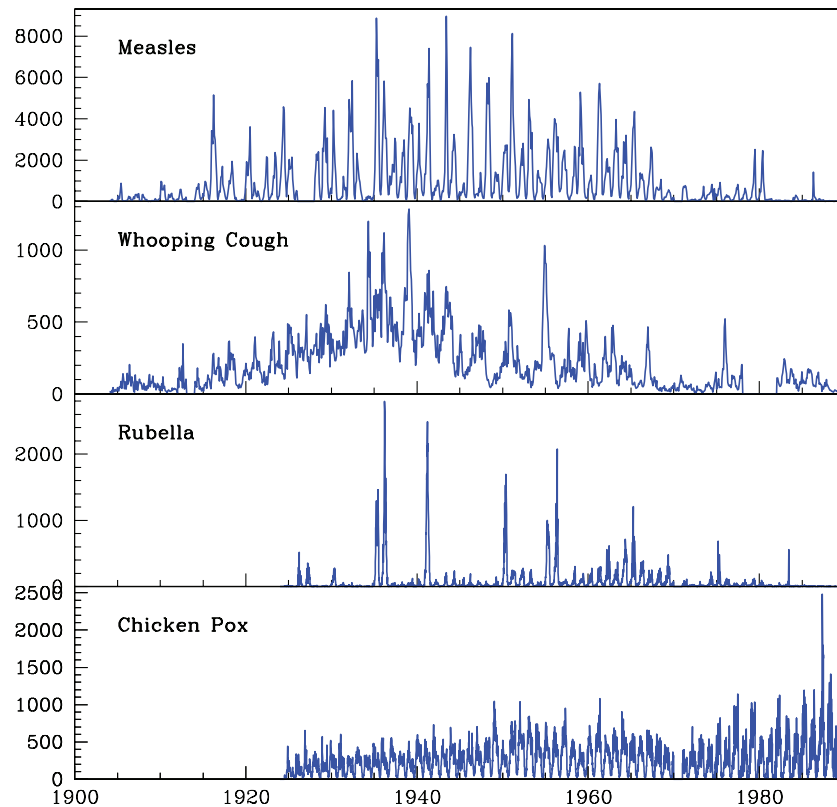

FIGURE 3. Reported cases of childhood infectious diseases in Ontario, Canada. For measles and whooping cough the data are monthly and begin in 1904. For rubella and chicken pox the data are weekly and begin in 1924.

2.3. Influenza

One of the diseases that we encounter most often is influenza, which is a major cause of morbidity in people of all ages and mortality in the elderly [**EDL02, DPV⁺06**]. The strong seasonality of influenza epidemics is clearly evident in the main panel, (a), of Figure 4, but this plot hides some critical features of influenza incidence patterns. The plot inset on the top right of panel (a) shows the first 20 years of data on a much larger vertical scale and emphasizes the true magnitude of the 1918 influenza pandemic. The other panels of this figure show that different indicators of the "influenza season" are consistent (panel b), that influenza epidemics occur in the winter in both the northern and southern hemispheres (panel c), that mortality data mask more complex patterns of incidence of different influenza types and subtypes (panel d), and that influenza evolves rapidly by a process known as *antigenic drift* (panel e) allowing the virus to evade our immune system again within a few years of infection.

None of the graphs in the figure reveal the most dangerous property of influenza, that every few decades (three times in the last century) it has undergone major changes which have left most of the world's human population suddenly susceptible to the newly invading strain. All three *antigenic shifts* in the last century (in 1918, 1957 and 1968) have originated from influenza viruses circulating in aquatic birds, which is why reports of "bird flu" infecting humans cause serious concern.

2.4. Statistical description of epidemic time series

All of the examples above have in common that epidemics of the disease in question are *recurrent*. If we wish to understand these epidemic patterns, then we must establish a precise way of quantifying the notion of "recurrence". Methods of *time series analysis* [**Cha04**] provide a standard set of tools to describe the temporal structure of recurrent epidemics.

The most natural way to begin analysis is with a *time plot*, like the graphs shown for each of the diseases discussed above. For a single epidemic, such as the Great Plague of London shown in top left panel of Figure 2, the time plot is an excellent descriptor. For recurrent epidemics, it is still the best starting point but is often difficult to interpret, especially if processes are occurring on a number of different scales. Figure 5 shows the original time series of cases of measles reported in England and Wales from 1944 to 1994 (top panel), together with two time plots of *transformed data*, which reveal much more structure (and noise) in the later parts of the time series when the overall number of cases had been greatly reduced as a result of mass vaccination.

Awareness of noise is important, but noise can mask important features that take place on longer timescales. For this reason, it is useful to make time plots of *smoothed data*, examples of which are shown in Figure 6. The plots shown here are of the England and Wales measles time series smoothed with a *moving average*, i.e., by replacing the original data points x_t with averages of nearby points,

$$(2.1) \qquad x_t \rightarrow \frac{1}{2a+1} \sum_{i=-a}^{a} x_{t+i} \, .$$
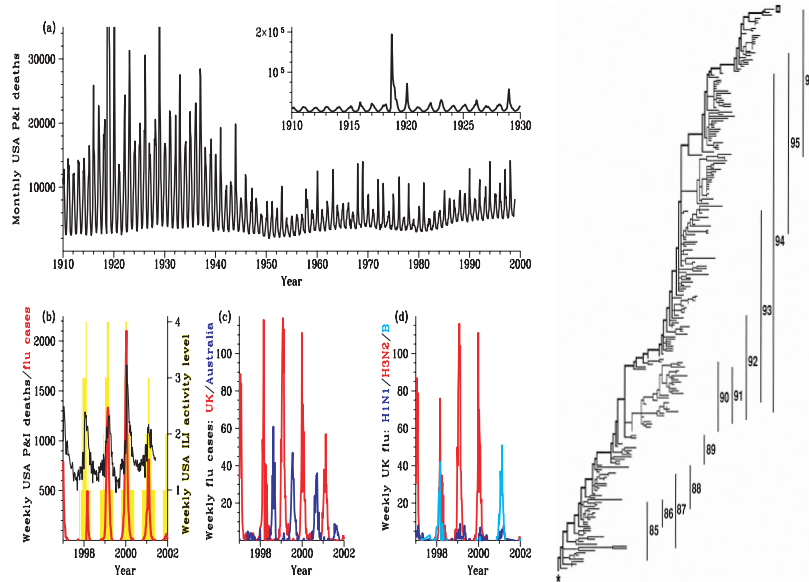
FIGURE 4. Influenza ecology and evolution. (a) Monthly mortality attributed to pneumonia and influenza (P&I deaths) in the USA in the 20th century. The inset plot shows the period 1910–1930 on a much larger scale, revealing the magnitude of the three peaks that extend beyond the top of the main panel: 1918–1919, 1919–1920 and 1928–1929. (b) Weekly measures of influenza incidence in the USA: P&I mortality in 122 cities (black curve), laboratory-confirmed cases (red curve), and influenza-like illness (ILI activity) on a five-point scale from 0 to 4 (yellow shading). (c) Weekly laboratory-confirmed cases in the UK (red curve) and Australia (blue curve). (d) Weekly patterns of relative prevalence of the three major groups of strains (H1N1, blue curve; H3N2, red curve; B, cyan curve) in the UK. (e) Maximum parsimony phylogenetic reconstruction of the evolution of the HA1 domain of the H3N2 subtype of influenza A from 1985 to 1997. The thick line running from the lower left (*=root) to the upper right (open square) is called the trunk and represents the successful H3N2 lineage. The vertical lines indicate the range of isolates from each flu year (1 October to 30 September). [Panels a–d reprinted from Earn et al [EDL02]; copyright © (2002), with permission of Elsevier. Panel e reprinted from Fitch et al [FBBC97]; copyright © (1997) National Academy of Sciences, USA.] Panel (d) gives a limited impression of the level of aggregation of data in the other panels. The coarsest data are those in panel (a), where the monthly records span periods during which different subtypes of influenza A have been present in human populations: 1918–1957 (H1N1), 1957–1968 (H2N2), 1968–1977 (H3N2), 1977–present (both H1N1 and H3N2). Data sources: US Vital Statistics [USVS], US CDC 122 cities data base (http://www.cdc.gov/epo/dphsi/121hist.htm), WHO FluNet data base (http://qamapserver.who.int/GlobalAtlas/)
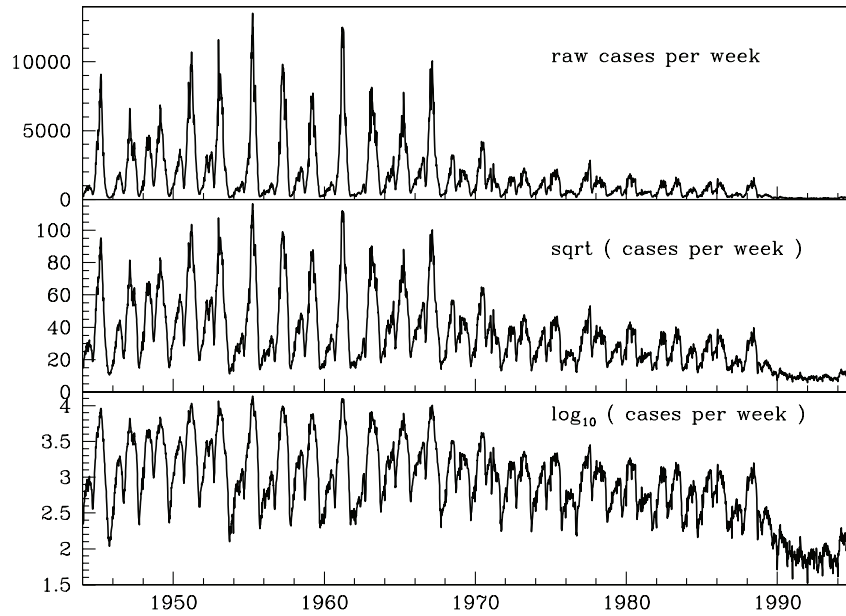
FIGURE 5. Transformed time plots of measles incidence in England and
Wales from 1944 to 1994.

A moving average is the simplest example of a *linear filter*,

$$(2.2) \qquad x_t \to \sum_{i=-\infty}^{\infty} \lambda_i x_{t+i} \,,$$

where the *weights* λ_i can be nonlinear functions of i [**Cha04**]. Linear filters are especially useful for extracting long-term trends that are clouded by noise or seasonality.

A useful way to think about recurrent patterns of epidemics is that the number of cases or deaths at a given time is correlated with the number at other times in the past or the future. As is probably familiar to you, given N pairs of observations of different quantities, $\{(x_i, y_i) : i = 1, \dots, N\}$, the *correlation coefficient* is defined to be

$$(2.3) \qquad r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}} \,,$$

where $\bar{x}$ and $\bar{y}$ are the means of $\{x_i\}$ and $\{y_i\}$, respectively. Given a single sequence of observations $\{x_t : t = 1, \dots, N\}$, we can compute the correlation of each observation with the observation k time steps in the future. Thus, we consider the pairs of observations $\{(x_t, x_{k+t}) : t = 1, \dots, N - k\}$ and define the *autocorrelation coefficient at lag k* to be

$$(2.4) \qquad r_k = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x}_{1,N-k})(x_{k+t} - \bar{x}_{k+1,N})}{\sqrt{\sum_{t=1}^{N-k}(x_t - \bar{x}_{1,N-k})^2 \sum_{t=1}^{N-k}(x_{k+t} - \bar{x}_{k+1,N})^2}} \,,$$
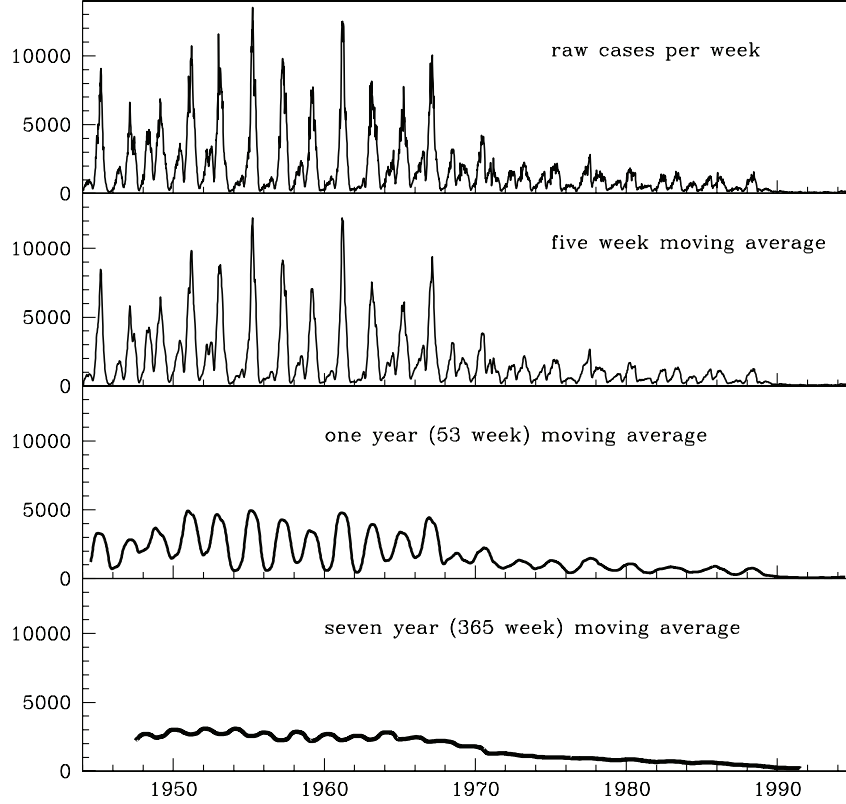
FIGURE 6. Smoothed time plots of measles incidence in England and
Wales from 1944 to 1994.

where $\bar{x}_{1,N-k}$ and $\bar{x}_{k+1,N}$ are the means of first and last $N-k$ observations, re-
spectively. If N is large and $k \ll N$ then $\bar{x}_{1,N-k} \simeq \bar{x}_{k+1,N} \simeq \bar{x}$, so

$$(2.5a) \qquad r_k \simeq \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{k+t} - \bar{x})}{\sqrt{\sum_{t=1}^{N-k}(x_t - \bar{x})^2 \sum_{t=1}^{N-k}(x_{k+t} - \bar{x})^2}}$$

$$(2.5b) \qquad \simeq \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{k+t} - \bar{x})}{\frac{N-k}{N}\sum_{t=1}^{N}(x_t - \bar{x})^2}.$$

Because this approximation depends on the assumption that $k \ll N$, the factor
$N/(N-k) \sim 1$ and is usually ignored, yielding the standard formula

$$(2.5c) \qquad r_k \simeq \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{k+t} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2}.$$

The factor $N/(N-k)$ increases with k, so ignoring it leads to an artificial reduction
in r_k as k increases, but if k is large enough that this could affect any inferences of
interest then a greater concern might be the approximation of the means of the first
and last $N-k$ observations by the mean of all observations. In general, it is worth
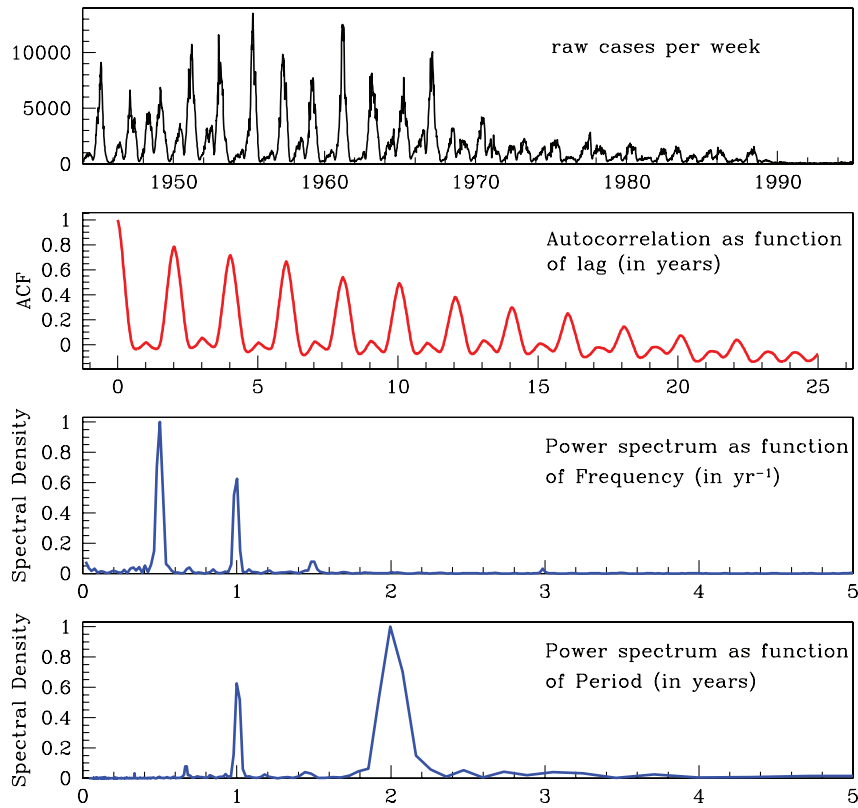
FIGURE 7. Time series analysis tools applied to measles incidence in England and Wales from 1944 to 1994. From top to bottom: the time plot, correlogram, periodogram, and period periodogram.

being aware that if the standard approximation (2.5c) is used then the precision of the autocorrelation coefficients decreases with lag.

A plot of the autocorrelation as a function of lag is known as the *correlogram* of the time series (see the second panel of Figure 7). The correlogram is a valuable indicator of the nature of the serial dependence in a time series. Nevertheless, since we are often more interested in the periodicities in the time series than in the degree to which correlation decays with lag, a frequency representation of the correlogram is usually preferred. The discrete Fourier transform of the correlogram is called the *periodogram*,

$$(2.6) \qquad I(\omega_p) = \frac{1}{\pi}\left(r_0 + 2\sum_{k=1}^{N-1} r_k \cos\omega_p k\right),$$

where $\omega_p = 2\pi p/N$, $p = 1, \ldots, N/2$. The periodogram can also be written directly in terms of the original time series. For example,

$$(2.7) \qquad I(\omega_p) = \frac{1}{\pi N \mathrm{var}[x_t]} \left|\sum_{t=1}^{N} x_t e^{i\omega_p t}\right|^2 .$$

You may wonder why this quantity is of greater interest than the Fourier transform of $\{x_t\}$. The answer is simply that it has a much more natural interpretation. Unlike the Fourier transform of $\{x_t\}$, the periodogram partitions the variance in $\{x_t\}$, i.e., $I(\omega_p)$ is the proportion of the variance associated with the harmonic $\cos \omega_p t$ (hence $\sum_{p=1}^{N/2} I(\omega_p) = 1$). Thus, while the positions of the peaks in the periodogram will coincide with those in the Fourier transform of $\{x_t\}$, the magnitude of these peaks is much easier to interpret in the periodogram.

In spite of its name, the periodogram is a function of frequency. It is often more convenient to plot it as a function of period, in which case it is called the *period periodogram*. Figure 7 shows the time plot, correlogram, periodogram and period periodogram for the measles case series in England and Wales. Note that while it is easiest to identify the *periods* of the peaks from the period periodogram, the area under the plotted curve is not straightforward to interpret. Area under the normal (frequency) periodogram for a range of frequencies is equal to the proportion of the variance of the time series associated with those frequencies.

Although the time series that we analyze contain discrete data points that are separated in time, the underlying processes occur in continuous time. We therefore think of the correlogram as an estimate of the underlying continuous *autocorrelation function* and of the periodogram as an estimate of the underlying (continuous) *power spectral density* or *power spectrum*, which is defined for all frequencies (all $\omega > 0$ rather than only $\omega_p = 2\pi p/N$ for $p = 1, \ldots, N/2$).

Because any observed time series that we analyze must be finite, the autocorrelation function is always less well sampled at longer lags (at lag k, the number of pairs of points being compared is $N - k$). Consequently, one might wish to put greater weight on the coefficients at shorter lags, which can be achieved by linearly filtering the autocorrelation coefficients in the calculation of the periodogram, i.e., by replacing Eq. (2.6) with

$$(2.8) \qquad I(\omega_p) = \frac{1}{\pi}\left(\lambda_0 r_0 + 2\sum_{k=1}^{M}\lambda_k r_k \cos \omega_p k\right).$$

Here, the set of weights $\{\lambda_k\}$ is called the *lag window* and M is known as the *truncation point*. Weighting coefficients "appropriately" is a bit of a black art, but a ubiquitous filter is the *Tukey window*,

$$(2.9) \qquad \lambda_k = \frac{1}{2}\left(1 + \cos\frac{\pi k}{M}\right), \qquad k = 0, 1, \ldots, M.$$

You can gain valuable intuition by applying this procedure to a time series of interest and looking at the effects of adjusting the truncation point (M). A variety of procedures for estimating the power spectrum of a time series are discussed and compared by Chatfield [**Cha04**].

A serious limitation of the periodogram is that it does not account for changes in the structure of a time series over time. If different segments of a time series correspond to different external conditions (such as different birth rates) then we must break up the time series ourselves before computing the periodogram. More sophisticated spectral methods that overcome this limitation do exist. *Wavelet analysis* [**TC98, Nie01, GBK01**] provides a method for frequency decomposition that is local in time, so you can see changes in the spectrum over time without having to identify distinct temporal segments yourself.

2.5. Exercises

The best way to learn the techniques described above is to try to apply them to an interesting time series.

(1) Download a time series of recurrent epidemics from the International Infectious Disease Data Archive [**IIDDA**]. Nice examples to try first are the monthly incidence of measles in New York City from 1928 to 1964 or the weekly incidence of measles in England and Wales from 1944 to 1994.

(2) Analyze the time series as a whole by constructing (i) time plots of raw and transformed data, (ii) correlograms, (iii) periodograms, (iv) period periodograms. Calculate the periodograms with and without a Tukey window, and examine the effects of different truncation points.

(3) Repeat these analyses for segments of the time series that look different by eye.

(4) Is there evidence for: Frequency components not evident by eye? Changes in the frequency structure over long time scales? If so, why might this have occurred? What if you remove trend and/or seasonality?

(5) If you're ambitious, explore the time series using wavelet analysis, e.g., via the MATLAB *Wavelet Toolbox* or Torrence and Compo's software [**TC98**] available at `http://atoc.colorado.edu/research/wavelets/`.

If you are so inclined, there are enough time series on the IIDDA web site to keep you busy making interesting discoveries for a long time. You can also find scripts on the IIDDA web site that will enable you to do all the analyses suggested in the exercises without programming anything yourself, but I encourage you to write your own code — you will learn more that way.

3. Modelling epidemics

As we have seen, the character of epidemic patterns differs for different diseases and even for the same disease in different places or at different times. Can models help us predict these differences, forecast future epidemics, or develop an understanding of the mechanisms responsible for the various patterns? Specifically, can models help us understand the structure of an epidemic time series as displayed by the time plot, correlogram or periodogram?

It is not clear at the outset how we should even begin trying to develop potentially useful models. Where you start will depend on what background and experience you have with other problems. I will discuss two standard approaches, which have their own strengths and weaknesses. The first approach is useful if you have little to go on beyond a time series of cases or deaths that you would like to understand, while the second approach is generally more powerful if you have more knowledge of how the fundamental biological processes might be related to observed events and some information in addition to a case or death time series.

3.1. Statistical modelling

Epidemic processes are ultimately stochastic. They depend on contacts between individuals (which occur randomly at least to some extent) and infections (which occur upon contact with some probability less than 1). We can build a *statistical model* using nothing but the fact there there is a stochastic process involved, and that the observed data we have are in the form of a time series $\{x_t\}$.

As a first attempt, we could imagine that the number of cases at time t, X_t, is simply a random variable with a known distribution. Our model for the epidemic process would then be simply

$$(3.1) \qquad\qquad X_t = \mu + Z_t \,,$$

where μ is the time average number of cases and Z_t is a sequence of independent and identically distributed (IID) random variables with zero mean. The parameters of this model are μ and the moments (of order 2 and higher) of the distribution of Z_t. We could then fit the model parameters to the mean, variance, etc. of our observed data. This might be a reasonable model for the importation of new, infectious individuals into a focal community, but it clearly fails to take into account that infection is transmitted from one individual to another, so there must be some correlation between the number of individuals in the focal community who are infected now and the number who will be infected in the near future.

So, imagine that the time series is generated by random processes, but that successive data points are correlated. For example, perhaps the data are generated by an *autoregressive* or "AR(p)" process,

$$(3.2) \qquad X_t - \mu = \alpha_1 (X_{t-1} - \mu) + \alpha_2 (X_{t-2} - \mu) + \cdots + \alpha_p (X_{t-p} - \mu) + Z_t \,,$$

where the α_i are constants that determine the degree of correlation along the time series and Z_t is as in Eq. (3.1). Alternatively, the data might be generated by a *moving average* or "MA(q)" process,

$$(3.3) \qquad\qquad X_t - \mu = \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q} \,,$$

where the β_i are constants that define a weighted average. More generally, the data might be generated by an *autoregressive moving average* or "ARMA(p,q)" process,

$$(3.4a) \qquad X_t - \mu = \alpha_1 (X_{t-1} - \mu) + \alpha_2 (X_{t-2} - \mu) + \cdots + \alpha_p (X_{t-p} - \mu)$$

$$(3.4b) \qquad\qquad\qquad + \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q} \,.$$

Of course, that the data *might* be generated by such a process is not especially enlightening. Can we identify a "good" model of our data from the infinite class of ARMA(p,q) processes?

You can develop some intuition about time series models by plotting the correlogram and periodogram for a variety of ARMA(p,q) processes with specific parameter values. The autocorrelation function of a first-order AR process declines exponentially, while the autocorrelation of an MA(q) process drops to zero at lag q. Unfortunately, for a more general ARMA(p,q) the theoretical autocorrelation function is complicated so, unless the correlogram of our observed time series is as simple as that of a pure AR(1) or MA(q) process, we need a quantitative method to identify the best ARMA(p,q) model.

But what do we mean by a "best" model? The usual answer is the most *parsimonious* model, i.e., the model that gets as close as possible to the data using the fewest parameters. If we think an observed time series $\{x_t\}$ was generated by an AR(p) process then we can attempt to find the best model by minimizing the residual sum of squares,

$$(3.5) \qquad S = \sum_{t=p+1}^{N} \left[(x_t - \mu) - \alpha_1 (x_{t-1} - \mu) - \cdots - \alpha_p (x_{t-p} - \mu) \right]^2 \,.$$

If we compute the least-squares fit for a range of orders (p) and plot S against p, we will likely find that there are diminishing returns for increasing the order beyond a certain point, which might suggest a "most parsimonious" AR(p) model. Of course, this model can be considered "best" only within the constraint that the distribution of noise (Z_t) is taken as given. Experience shows, moreover, that fewer parameters are typically required to model a process as ARMA(p, q) rather than pure AR(p) or MA(q) processes. So we really need to compute the residual sum of squares for ranges of p and q and for a variety of possible noise distributions, and consider the number of parameters involved in each model. This is not straightforward. We really need a quantitative criterion that we can apply blindly to a wide variety of models. Unsurprisingly, the development of such a criterion is something to which statisticians have devoted much time and effort.

We know that statistical models involving more parameters will generally yield fits that are closer — in the least squares sense — to observed data. We need a formal away to penalize models for using more parameters so we can arrive at an objective measure of the statistical significance of the fit. There are many ways to do this, the most common being the *Akaike Information Criterion* (AIC). The AIC depends on the notion of the *likelihood* of the model given the data, which is defined to be the probability of observing the data given the model. If $L(\theta|x_t)$ denotes the likelihood of the model with parameter values θ, given the observed time series $\{x_t\}$, then the AIC is

$$(3.6) \qquad \qquad \mathrm{AIC} = -2\ln\left[\max_\theta L(\theta|x_t)\right] + 2k\,,$$

where k is the number of independent parameters that are fitted for the model being assessed (i.e., the dimension of the vector θ). By choosing the model with the lowest value of the AIC we balance the closeness of the fit (maximum likelihood) with the complexity of the model (minimum number of parameters). It is often assumed that measurement errors in the data are normally distributed about the "true" model, in which case maximizing the likelihood is equivalent to minimizing the residual sum of squares [**PTVF92**] (e.g., Eq. 3.5 for an AR(p) model). With this assumption, Eq. (3.6) becomes

$$(3.7) \qquad \qquad \mathrm{AIC} = N\ln(S/N) + 2k\,,$$

where N is the number of observations in the time series.

Once we have arrived at a "best fit" model, we can use it to *forecast* future observations. The natural way to test this procedure is to fit a model to part of the observed time series and then use the resulting fit to predict the remainder of the data. How successful is this likely to be for an infectious disease time series like one of those discussed above?

The answer depends on which time series we are interested in and which segment of the data we start with. Consider, for example, weekly chicken pox in Ontario (Figure 3). If we fit an ARMA(p, q) model using the first 20 years of the 50 year time series and then use the model to forecast the remaining 30 years of data, we can expect to do reasonably well. On the other hand, if we fit an ARMA(p, q) model to the New York City monthly measles reports from 1928 to 1945 (top right panel of Figure 2), we will fail miserably at predicting the monthly case reports from 1946 to 1964. The reason is that the first and second halves of the New York City measles time series have completely different structure (apparently irregular until 1945 and strictly biennial after 1945). We cannot expect to predict the second

half given only the first half. We need further information — such as the fact that the birth rate rose dramatically after 1945 — and we need to know how to alter our fitted ARMA(p, q) model in light of this information. Unfortunately, we will never have the information required to alter the fitted parameters of our statistical time series model, because the parameters of the model have no biological meaning. This is a serious problem that is not easy to overcome.

You might hope that you could do better with a more sophisticated time series model. ARMA(p, q) models are relatively simple and you can often obtain better fits to data (with lower AIC) using more elaborate statistical models [**Scu03**]. In particular, a time series analyst would probably advise you to first fit trends and cycles to the observed data and remove these features before attempting to fit an ARMA(p, q) or more sophisticated model. But this emphasizes the problem, because we are usually more interested in understanding trends and changes in cyclicity than in modelling small stochastic fluctuations in detail.

Ultimately, statistical time series models are useful as forecasting tools only if the underlying stochastic processes are *stationary*, which means roughly speaking that the processes are identical for the period during which the parameters are fit and the period over which predictions are sought. If external conditions change in a way that affects the processes that generate the data we observe, then we need a different modelling tool.

Much more detail on statistical modelling of time series can be found in the excellent text by Chatfield [**Cha04**].

3.2. Mechanistic modelling

A fundamental problem with statistical time series models is that they use the observed time series itself to estimate parameters. It would be much better to have a model that we can parameterize from independently collected data and then explore whether that model can explain the observed time series. This more desirable approach can be pursued only if we have a lot more information than the time series we are trying to explain. Fortunately, when studying time series of infectious disease epidemics this is precisely the situation in which we usually find ourselves.

Rather than beginning with a time series, let's consider what we know about the biological processes that must ultimately give rise to the observed cases or deaths, i.e., the mechanisms of disease transmission and spread. To be specific, think about a childhood disease such as measles or chicken pox. Initially, perhaps after a period of maternally-acquired immunity has waned, each individual is *susceptible* to the disease in question. If some individuals in the population have been exposed to the disease-causing agent and are currently at the *infectious* stage then they can transmit the pathogen to susceptible individuals. For typical childhood diseases, recovery entails essentially lifelong immunity, so regardless of whether infected individuals eventually recover or die, we can consider them to be *removed* from the transmission process.

We are thus naturally led to a simple model in which the host population (humans) is divided into three compartments, each containing individuals who are identical in terms of their status with respect to the disease in question. In this "SIR model", which is depicted as a flow chart in Figure 8, the three compartments are:

- *Susceptible*: individuals who have no immunity to the infectious agent, so might become infected if exposed.
- *Infectious*: individuals who are currently infected and can transmit the infection to susceptible individuals who they contact.
- *Removed*: individuals who are immune to the infection, and consequently do not affect the transmission dynamics in any way when they contact other individuals.

It is traditional to denote the number of individuals in each of these compartments as S, I and R, respectively. The total host population size is $N = S + I + R$.
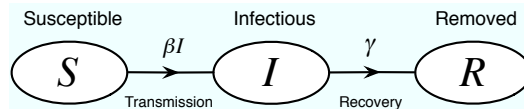

FIGURE 8. The basic SIR model. Individuals in the host population are in one of three possible states: *Susceptible*, *Infectious*, or *Removed*. The rates of flow between compartments are specified relative to the number of individuals in the state in question. The differential equation version of this model is given in Eq. (3.8).

Having compartmentalized the host population, we now need a set of equations that specify how the sizes of the compartments change over time. Solutions of these equations will yield, in particular, $I(t)$, the size of the infectious compartment at time t. A plot of $I(t)$ should bear a strong resemblance to observed epidemic curves if this is a reasonable model.

The numbers of individuals in each compartment must be integers, of course, but if the host population size N is sufficiently large we can treat S, I and R as continuous variables and express our model for how they change in terms of a system of differential equations,

$$(3.8a) \qquad \frac{dS}{dt} = -\beta SI \,,$$

$$(3.8b) \qquad \frac{dI}{dt} = \beta SI - \gamma I \,.$$

Here, the *transmission rate* (per capita) is β and the *recovery rate* is γ (so the *mean infectious period* is $1/\gamma$). In writing Eqs. (3.8) we are assuming implicitly that contacts among individuals are random and, moreover, that the population is homogeneously mixed (like an ideal gas). Note that we do not need a differential equation for the number of removed individuals. The appropriate equation is $dR/dt = \gamma I$ (outflow from the I compartment goes into the R compartment) but since R does not appear in equations (3.8) the equation for dR/dt has no effect on the dynamics of S and I (formalizing the fact that removed individuals cannot affect transmission). This basic SIR model has a long history [**KM27**] and is now so standard that you can even find it discussed in some introductory calculus textbooks [**HHGL⁺06**].

If everyone is initially susceptible ($S(0) = N$), then a newly introduced infected individual can be expected to infect other people at the rate βN during the expected infectious period $1/\gamma$. Thus, this first infective individual can be expected to infect $\mathcal{R}_0 = \beta N/\gamma$ individuals. The number $\mathcal{R}_0$ is called the *basic reproduction number*

and is unquestionably the most important quantity to consider when analyzing any epidemic model for an infectious disease. In particular, $\mathcal{R}_0$ determines whether an epidemic can occur at all; to see this for the basic SIR model, note in equations (3.8) that I can never increase unless $\mathcal{R}_0 > 1$. This makes intuitive sense, since if each individual transmits the infection to an average of less than one individual then the number of cases must decrease with time.

Intuitively, if you introduce a pathogen with $\mathcal{R}_0 > 1$ into a fully susceptible population then you expect the number of cases to grow exponentially until a large fraction of the susceptible have been infected. The epidemic curve must eventually turn over. Simple analyses of the basic SIR equations (3.8) with $\mathcal{R}_0 > 1$ show that the SIR model does exactly this (see exercises below). But it is also easy to show that Eqs. (3.8) always lead to exactly one maximum in $I(t)$ so we can never predict more than one epidemic. This means that the basic SIR model cannot account for the recurrent epidemics described above. In fact, it can't even explain persistence of a disease in a population. We must include more structure in our model, which we can do by considering more carefully the biological processes involved.

One aspect of disease transmission that we have ignored is that there is a delay between the moment at which an individual is infected and the beginning the period of infectiousness (the *latent period*). To account for this delay, an *exposed* compartment is usually added ("exposed" individuals are infected but not yet infectious), yielding an "SEIR model". In this way, we introduce one more equation and one more parameter, the rate (σ) of flow from the exposed to the infectious compartment. The mean latent period ($1/\sigma$) can often be estimated, so adding this parameter doesn't get us into trouble, but the dynamics of the basic SEIR model are almost identical to the SIR model. We still cannot explain persistence, let alone cyclical recurrence. We are still missing something fundamental.

The key element that is absent from our discussion so far is a source of susceptible individuals, the most obvious being births. If we are considering timescales over which births are important then we also need to account for deaths (from causes other than the disease in question). The streams of births and deaths together are known as the *vital dynamics* of the system. Of course, only births of individuals who are not immune affect the transmission dynamics of the disease in question. For the childhood diseases I have mentioned, many young children are vaccinated before ever being exposed to the associated pathogens, so we can consider the proportion of the population who has been vaccinated to have been born into the removed compartment. If the birth rate is ν ("nu for natality"), the *per capita* death rate is μ ("mu for mortality") and the proportion vaccinated is p, then we arrive at the "SEIR model with vital dynamics and vaccination" depicted as a flow chart in Figure 9. Expressing this model using differential equations we obtain

$$(3.9a) \qquad \frac{dS}{dt} = \nu(1-p) - \beta SI - \mu S \,,$$

$$(3.9b) \qquad \frac{dE}{dt} = \beta SI - \sigma E - \mu E \,,$$

$$(3.9c) \qquad \frac{dI}{dt} = \sigma E - \gamma I - \mu I \,,$$

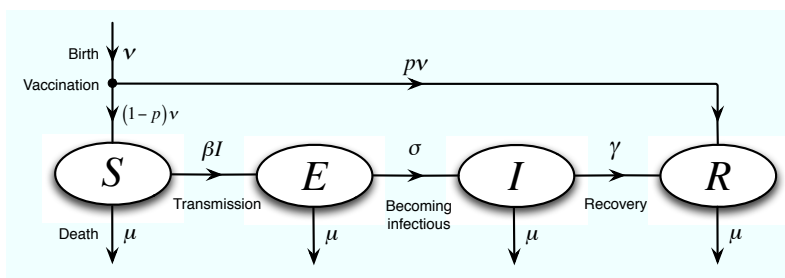$$(3.9d) \qquad \frac{dR}{dt} = \nu p + \gamma I - \mu R \,.$$

FIGURE 9. The SEIR model with vital dynamics and vaccination. This expands the model depicted in Figure 8 to include a class of individuals who have been *Exposed* to the pathogen and are currently infected, but are not yet infectious. In addition, births occur at rate ν, deaths occur at *per capita* rate μ, and the rate at which exposed individuals become infectious is σ (so $1/\sigma$ is the *mean latent period*, the mean time from the moment of initial infection to the beginning of the infectious period). This model is represented as a system of differential equations in Eq. (3.9).

If births exactly balance deaths ($\nu = \mu N$) then the population size ($N = S+E+I+R$) is constant and one of these equations is superfluous. In this expanded model, the basic reproduction number $\mathcal{R}_0$ is still approximately, but no longer exactly, $\beta N/\gamma$ (see exercises below).

Equilibrium and stability analyses of Eqs. (3.9) lead to important conclusions. For any parameter set, there is always at least one equilibrium, the *disease free equilibrium* or DFE corresponding to no infected individuals at all ($E = I = 0$). If $\mathcal{R}_0(1-p) < 1$ then the DFE is globally asymptotically stable. If $\mathcal{R}_0(1-p) > 1$ then the DFE is unstable and there is a second equilibrium, the *endemic equilibrium* corresponding to a constant number of exposed and infected individuals ($E > 0$ and $I > 0$), which is globally asymptotically stable [**LM95, KM04**].

The existence of a globally stable endemic equilibrium immediately provides an explanation for persistence of infectious diseases. However, it is the condition for global stability of the DFE that is in fact most important. Since the proportion vaccinated is under our control, we can re-arrange the condition $\mathcal{R}_0(1-p) < 1$ to find a critical vaccination proportion,

$$(3.10) \qquad\qquad p_{\text{crit}} = 1 - \frac{1}{\mathcal{R}_0},$$

above which the disease cannot persist. Consequently, if we can estimate $\mathcal{R}_0$ for a given disease then we can also estimate the threshold vaccination level that we must reach in order to eradicate the infection. This is probably the single most important conclusion that has been drawn from mathematical analysis of epidemic models, and it helps us understand why smallpox (with $\mathcal{R}_0 \sim 4$) was successfully eradicated in the 1970s, while measles (with $\mathcal{R}_0 \sim 20$) is still with us. But from the point of view of our present discussion, the critical vaccination level is just a bonus. We were actually aiming to understand something else, the recurrent pattern of epidemics seen in most of the data we discussed.

A standard local stability analysis (see exercises below) reveals that the endemic equilibrium is normally approached by damped oscillations with period

$$(3.11) \qquad T = 2\pi \sqrt{\frac{1/\sigma + 1/\gamma}{\nu(\mathcal{R}_0 - 1)}}$$

This suggests an explanation of cyclical epidemic patterns and a method to predict the inter-epidemic period. However, as the examples discussed above show, epidemic cycles seen in practice are often undamped, and the inter-epidemic period can change over time and differ from place to place. So, what are we missing?

3.3. Demographic stochasticity

One thing we have glossed over is the presence of noise. While it is true that for sufficiently large population size N it is reasonable to treat S, E, I and R as continuous variables, it is not true that the genuine discreteness of the number of individuals in each compartment has no observable effect.

If we re-cast the SEIR model as a stochastic process [**Bai75**] then our ability to derive conclusions analytically rapidly evaporates. Fortunately, we can simulate stochastic models on a computer and numerically explore the significance of demographic noise. Gillespie [**Gil76**] showed that a simple algorithm provides rigorously correct realizations of the sorts of stochastic processes that arise from systems like the SIR and SEIR models that concern us here. So, we can run many realizations and infer the expected behaviour of genuinely stochastic epidemics.

Gillespie's algorithm works as follows. Let a_1, a_2, ... , be the rates at which the various processes occur. For example, a_1 might be the birth rate, a_2 might be the rate of going from susceptible to exposed, a_3 might be the rate of going from infectious to removed (recovering) and so on. Let a_0 be the overall event rate, i.e., $a_0 = \sum_i a_i$, so the average time between events is $1/a_0$. If the transitions between the various states are Poisson processes then the time spent in any given state is exponentially distributed and the probability that the next event occurs between times t and $t + dt$ is $a_0 e^{-a_0 t} dt$. If we let $u = e^{-a_0 t}$ then $u \in [0,1]$ and $du = a_0 e^{-a_0 t} dt$, so u is uniformly distributed in $[0,1]$. Thus, the time t to the next event can be obtained by sampling u from a uniform distribution in $[0,1]$ and setting $t = \frac{1}{a_0} \ln(1/u)$. The nature of the event is then determined by sampling a point from a uniform distribution on $[0, a_0]$ and setting the event to type i if the uniform deviate lies in the ith interval in the following list.

$$(3.12) \qquad [0, a_1), [a_1, a_1 + a_2), \dots , [a_1 + \cdots + a_{i-1}, a_1 + \cdots + a_i), \dots$$

Using Gillespie's algorithm, we can easily confirm the discovery made 50 years ago by Bartlett [**Bar57**] that a relatively small amount of noise is sufficient to prevent the oscillations of the SIR model from damping out. In the stochastic SIR model [**Bai75, Gil76**] demographic noise sustains the oscillations that damp out in the deterministic model (Figure 10).

Again, this is progress that has arisen from an important mechanistic insight. But we are left with another puzzle. If you look carefully at the New York City measles reports in the bottom right panel of Figure 2 you'll see that before about 1945 the epidemics were fairly irregular, whereas after 1945 they followed an almost perfect two-year cycle. Even with oscillations sustained by noise, the SIR model cannot explain why the measles epidemic pattern in New York City changed in this way. Have we missed another important mechanism?
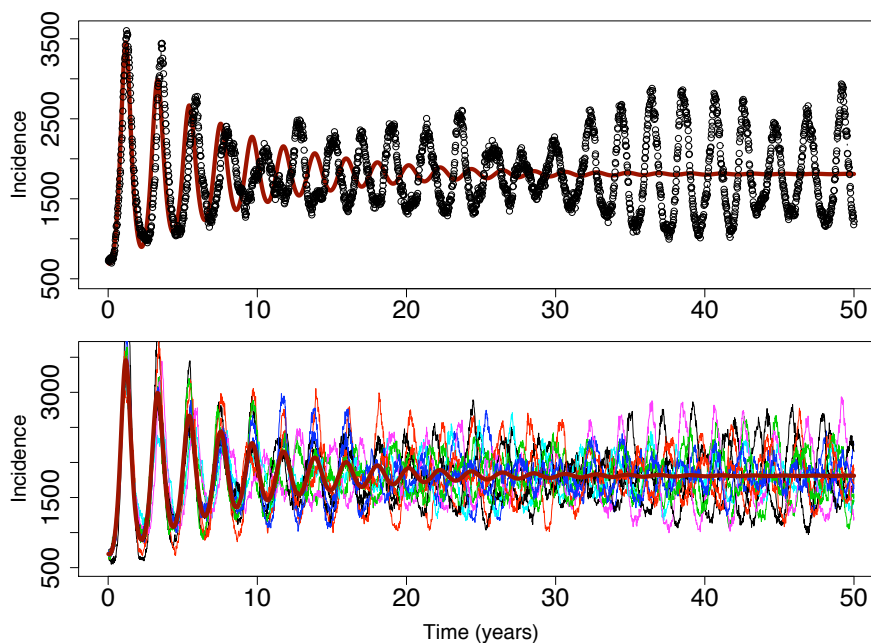
FIGURE 10. Simulations of the SEIR model model with vital dynamics, for a population of 5 million individuals. The birth and death rates are $\nu = \mu = 0.02$, the basic reproduction number is $\mathcal{R}_0 = 17$, and the mean latent and infectious periods are 8 and 5 days, respectively. In both panels, the heavy solid red line is the solution of the deterministic model (3.9). The top panel also shows a single realization of the associated stochastic process computed with the Gillespie algorithm; it is clear that a small amount of noise prevents the deterministically damped oscillations from decaying. The bottom panel shows 10 Gillespie realizations starting from the same initial conditions.

3.4. Seasonal forcing

So far, we have been assuming implicitly that the transmission rate β (or, equivalently, the basic reproduction number $\mathcal{R}_0$) is simply a constant and, in particular, that it does not change in time. Let's think about that assumption. The transmission rate is really the product of the rate of contact among individuals and the probability that a susceptible individual who is contacted by an infectious individual will become infected. But the contact rate is *not* constant throughout the year. To see that, consider the fact that in the absence of vaccination the average age at which a person is infected with measles is about five years [**AM91**], hence most susceptibles are children. Children are in closer contact when school is in session, so the transmission rate must vary seasonally.

A crude approximation of this seasonality is to assume that β varies sinusoidally,

$$(3.13) \qquad \beta(t) = \langle\beta\rangle(1 + \alpha \cos 2\pi t).$$

Here, $\langle\beta\rangle$ is the mean transmission rate, α is the *amplitude* of seasonal variation $(0 \le \alpha \le 1)$ and time t is assumed to be measured in years. If, as above, β is

assumed to be a periodic function (with period one year) then the SIR model is said to be *seasonally forced*.

You might think that seasonal forcing is just a minor tweak of the model. In fact, this forcing has an enormous impact on the epidemic dynamics that the model predicts. If you've ever studied the forced pendulum then you might already have some intuition for this. A pendulum with some friction will exhibit damped oscillations and settle down to an equilibrium. But if you tap the pendulum with a hammer periodically then it will never settle down and it can exhibit quite an exotic range of behaviours including chaotic dynamics [**Gle87, Str94**] (oscillations that look random). Similarly complex dynamics can occur in the seasonally forced SIR model.

Importantly, with seasonal forcing the SIR model displays *un*damped oscillations, even in the absense of stochasticity. More than that, for different parameter values, the seasonally forced SIR model can produce all the different types of oscillatory measles patterns that have ever been seen in real data. The range of possible dynamical behaviour is summarized by the *bifurcation diagram* shown in Figure 11.

We are definitely making good progress. With a fixed set of parameters we can explain epidemic cycles of many different periods (whereas, for given set of parameters, the unforced model always predicted the same oscillation period given by Eq. (3.11)). However, as noted in the previous section, the observed measles epidemics in New York City show very clearly that the dynamical pattern changed over time (bottom right panel of Figure 2) and other significant qualitative changes have been observed in measles case series in other places [**ERBG00**]. How can we explain *changes over time* in the *pattern* of epidemics?

If you are familiar with chaotic dynamics, then one possible solution might occur to you, namely *intermittency* [**PST93**]. Some chaotic dynamical systems appear to switch randomly from one type of oscillation to another. Perhaps transitions from cycles of various kinds to irregular epidemics are manifestations of intermittency in a system that is fundamentally chaotic? It has been proposed that measles dynamics are chaotic [**SK85, OS90, SKTO93**], but there is a serious problem with the chaotic solutions of the seasonally forced SIR model: as is evident from Figure 11, they often reach absurdly small infected populations (e.g., proportion infected $< 10^{-15}$). In practice, this means that if the epidemic pattern is chaotic then the pathogen can be expected to go extinct in a relative short time, which is not observed.

What biological mechanism that we have not yet included might repair the incorrect prediction of rapid extinction? One possibility might be that the population is not really homogeneously mixed. In particular, young children mix a lot more with others of their own age than they do with teenagers or adults. The influence of age-dependent contact structures on measles dynamics was investigated by Schenzle [**Sch84**], who developed a *realistically age-structured* (RAS) model involving 84 differential equations (SEIR in each of 21 age classes). Analysis of this model showed that it did seem to avoid predicting rapid extinctions and that it seemed to explain the detailed temporal dynamics of individual measles time series [**BG93**].

However, later analysis indicated that the critical ingredient in the RAS model was not really that contact rates are age structured, but that they are higher during school terms [**ERBG00**]. This may seem like a step backwards — after all, the point of introducing seasonal forcing was precisely to model the seasonal variation
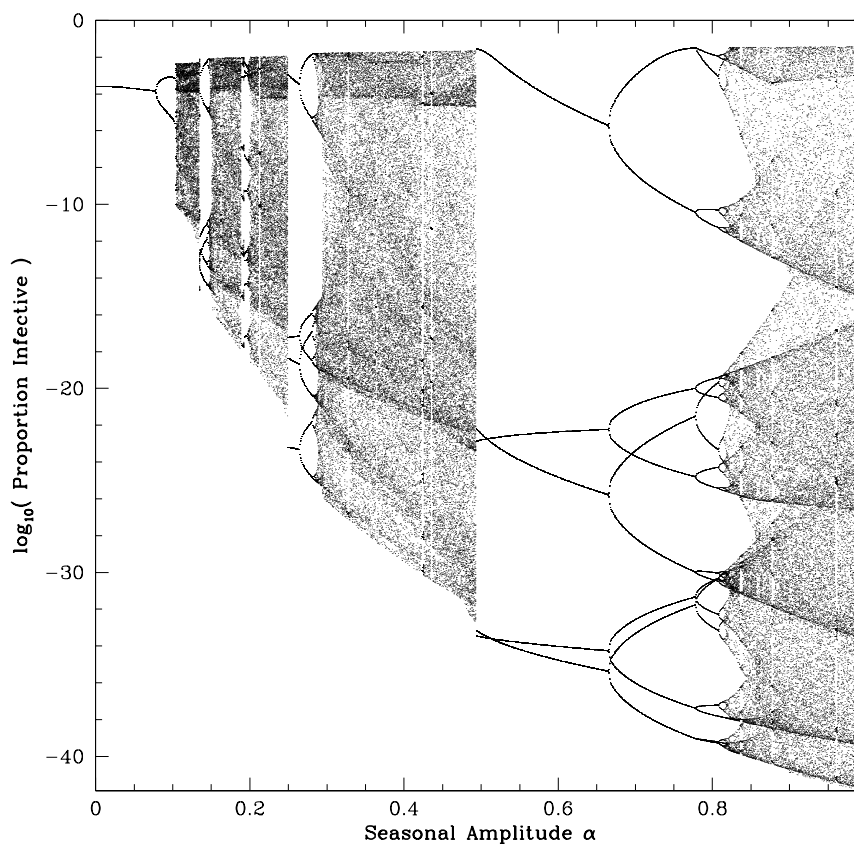
FIGURE 11. Bifurcation diagram for the sinusoidally forced SIR model with $\mathcal{R}_0 = 18$, $1/\gamma = 5$ days, $\mu = 0.02\,\mathrm{yr}^{-1}$. The seasonal amplitude α is defined by Eq. (3.13). For each value of α, the sinusoidally forced SIR equations were integrated using the standard fourth order Runge-Kutta scheme [**PTVF92**] with a time-step of 0.5 day. After ignoring a 1000 year transient, the proportion of the population that was infectious (I) was plotted once a year for 200 years. Thus, a single point above a given value of α indicates an annual cycle, two points correspond to a biennial cycle, etc., and a broad band of points indicates aperiodic behaviour (chaos). Because only one solution is plotted for each α, this bifurcation diagram is incomplete. For some ranges of α there are, in fact, multiple co-existing stable cycles [**SS83**].

in contact rates due to school terms. But the true variation in contact rates is not sinusoidal. Schenzle's RAS model includes not only 21 age classes, but genuine *term-time forcing*, i.e., rather than Eq. (3.13), the RAS model uses

$$(3.14) \qquad \beta(t) = \begin{cases} \beta_{\mathrm{H}} & \text{school days,} \\ \beta_{\mathrm{L}} & \text{non-school days.} \end{cases}$$

i.e., the transmission rate is high on days when school is in session (β_{H}) and low otherwise (β_{L}). For the purposes of comparing results with those obtained with the sinusoidally forced model, it is more convenient to express the term-time forcing function in terms of the mean transmission rate $\langle\beta\rangle$ and the amplitude of seasonal forcing α. If p_{s} is the proportion of days in school then the mean transmission rate is

$$(3.15) \qquad\qquad \langle\beta\rangle = p_{\mathrm{s}}\beta_{\mathrm{H}} + (1 - p_{\mathrm{s}})\beta_{\mathrm{L}}$$

and we define the amplitude of seasonality to be

$$(3.16) \qquad\qquad \alpha = \frac{1}{2}\left(\frac{\beta_{\mathrm{H}} - \beta_{\mathrm{L}}}{\langle\beta\rangle}\right).$$

In terms of $\langle\beta\rangle$ and α we can write the transmission rate

$$(3.17) \qquad \beta(t) = \begin{cases} \left[(1 + 2(1 - p_{\mathrm{s}})\alpha\right]\langle\beta\rangle & \text{school days,} \\ \left[1 - 2p_{\mathrm{s}}\alpha\right]\langle\beta\rangle & \text{non-school days.} \end{cases}$$

This form for $\beta(t)$ is more convenient than (3.14) because it is $\langle\beta\rangle$ that corresponds to β in the unforced model (for which $\alpha = 0$). Note that since $\beta(t) \geq 0$, it is clear from equation (3.17) that the range over which the seasonal amplitude α can be varied is

$$(3.18) \qquad\qquad 0 \leq \alpha \leq \frac{1}{2p_{\mathrm{s}}}.$$

At the extreme of $\alpha = 1/2p_{\mathrm{s}}$, $\beta_{\mathrm{L}} = 0$, i.e., $\beta(t) = 0$ on non-school days. As an example, note that if we apply Eq. (3.18) to England and Wales, where a typical year contains 198 school days, we must have $0 \leq \alpha \leq 365/396 \simeq 0.9217$.

In terms of its ability to account for observed measles dynamics, the term-time forced SEIR model is just as successful as the full RAS model, and avoids the need to estimate—or guess the values of—many new parameters [**ERBG00**]. The model predicts that the disease will persist and it reproduces many detailed features of the biennial pattern of measles epidemics observed in New York City and London. But it does perhaps *too good* a job of reproducing this behaviour! No other epidemic patterns are predicted, so we are back to a situation in which we cannot explain changes in the cyclical structure of the time series. We will need to think again if we want to predict the observed dynamical transitions.

3.5. Exercises

If you are not familiar with the notions of local and global asymptotic stability, or you need to refresh your memory, you can consult any standard text on dynamical systems (e.g., [**GH83, Str94, Wig03**]) or mathematical ecology (e.g., [**BCC01**]).

(1) Find an exact analytical formula for the phase portrait of solutions to the basic SIR equations (3.8). *Hint:* Consider dI/dS.

(2) In the basic SIR model (3.8), show that the number of cases must eventually decrease, regardless of initial conditions or parameter values.

(3) Consider the SIR and SEIR models with vital dynamics, but without vaccination.
   (a) Use biological reasoning to derive an exact expression for $\mathcal{R}_0$ and show that $\mathcal{R}_0 \simeq \beta N/\gamma$. *Hint:* Consider the average length of time spent in each compartment.

(b) Prove that there is always a disease free equilibrium (DFE) and there is an endemic equilibrium if and only if $\mathcal{R}_0 > 1$.

(c) Show that the DFE is globally asymptotically stable if $\mathcal{R}_0 < 1$ and unstable if $\mathcal{R}_0 > 1$. Show that the endemic equilibrium is locally asymptotically stable if $\mathcal{R}_0 > 1$, and that for parameter values that could correspond to a childhood disease the equilibrium is approached by damped oscillations (the relevant eigenvalue has a non-zero imaginary part) with period given by Eq. (3.11).

(4) Use a numerical ODE solver to explore solutions of the sinusoidally forced SIR model for measles ($\mathcal{R}_0 = 18$, mean infectious period $1/\gamma = 5$ days) for a range of seasonal forcing amplitudes. Compare the character of the solution with what you would have guessed from the bifurcation diagram in Figure 11.

4. Predicting epidemics

We have converged on a model—the term-time forced SEIR model—that does an excellent job of reproducing measles dynamics in a few places at particular times (e.g., New York City from 1945 to 1964, bottom right panel of Figure 2). To explain the occurrence of different measles patterns in other places and at other times, we do not need to include more structure in our model. We simply need to recognize that we have assumed implicitly that all the parameters of the model are fixed constants, whereas some of these parameters might change over time for reasons that are external to the process of disease transmission. In particular, the birth rate (ν) is not constant, and neither is the proportion of the population that is vaccinated (p).

We might, therefore, be able to predict different dynamics based on different birth rates and vaccination levels. Before attempting to do so, however, it is extremely useful to note [**ERBG00**] that changes in ν or p are equivalent—in terms of their dynamical influence—to changes in the mean transmission rate $\langle \beta \rangle$ (or, alternatively, the basic reproduction number $\mathcal{R}_0$). To see this, suppose that a proportion of newborns is vaccinated (after maternally acquired immunity wanes) so $p > 0$ in Eqs. (3.9a) and (3.9d). Now consider a simple change of variables,

(4.1)
$$S = S'(1-p), \qquad E = E'(1-p), \qquad I = I'(1-p), \qquad R = R'(1-p) + \frac{\nu}{\mu}p.$$

The dynamical equations for the primed variables are exactly the SEIR equations (3.9) *without vaccination*, except that the transmission rate β is everywhere replaced by $\beta(1-p)$. Thus, except for an overall reduction in the number of cases, when a proportion of a population is vaccinated the dynamics are identical to those of an unvaccinated SEIR system with a smaller mean transmission rate, $\langle \beta \rangle \rightarrow \langle \beta \rangle(1-p)$.

Since the birth rate ν enters Eq. (3.9a) in exactly the same place as vaccination, a similar argument applies to changes in birth rate. A change in birth rate from ν to ν' is dynamically equivalent to a change in mean transmission rate from $\langle \beta \rangle$ to $\langle \beta \rangle(\nu'/\nu)$. More generally, if the birth rate changes from ν to ν' and the vaccination proportion changes from p to p' then the resulting dynamics are identical to those

based on no change in ν or p but a change in the mean transmission rate,

$$(4.2) \qquad \langle\beta\rangle \to \langle\beta\rangle \frac{(1-p')}{(1-p)} \frac{\nu'}{\nu} \,.$$

This strongly suggests that we will be able to predict changes in dynamics based on demographic data (which tell us how ν and p change with time) and a single bifurcation diagram, where the control parameter is the mean transmission rate $\langle\beta\rangle$. The predicted dynamics will depend on the *susceptible recruitment rate*, $\nu(1-p)$.
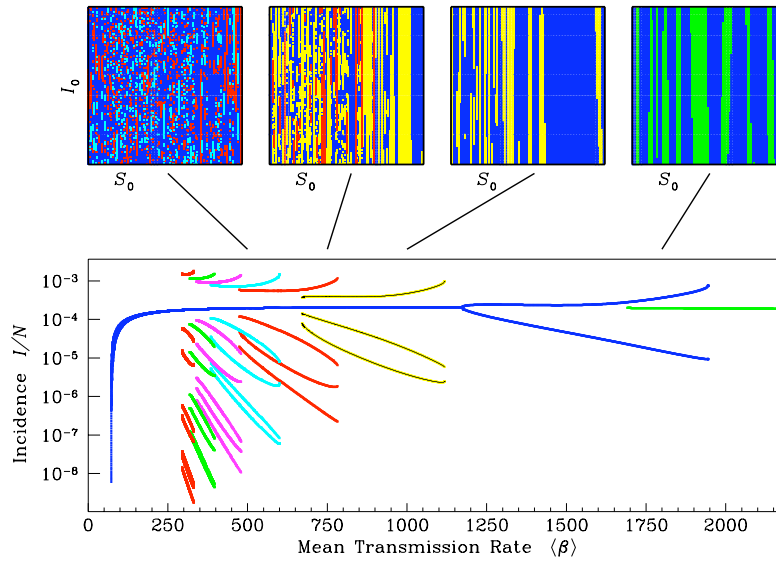

FIGURE 12. The main panel is the bifurcation diagram for the term-time forced SEIR model, showing incidence on 1 January, normalized by (constant) population size; the control parameter is the mean transmission rate $\langle\beta\rangle$. The fixed parameter values are $1/\gamma = 5$ days, $1/\sigma = 8$ days, $\mu = \nu/N = 0.02\,\mathrm{yr}^{-1}$ and $\alpha = 0.25$ (Eq. (3.17)). Each attractor is identified with a different color. For sufficiently high $\langle\beta\rangle$, there is a unique (annual) attractor. As $\langle\beta\rangle$ is reduced, biennial, 3-, 4-, 5-, 6-, 7-, and 8-year cycles all occur, before all but the annual attractor are extinguished. The term-time forcing function used to produce this figure corresponds to school terms in England and Wales [**BE03**]. In the United States, the summer holiday is longer; this does not affect the structure of the bifurcation diagram, but with a longer summer holiday each of the various bifurcations occurs at lower $\langle\beta\rangle$. Above the bifurcation diagram, basins of attraction (initial susceptibles, $0 < S_0/N < 0.1$, versus initial infectives, $0 < I_0/N < 0.0001$, with $E_0/N = 0.0001$) are shown for the various attractors at four particular values of $\langle\beta\rangle$. Figure 13 identifies regions of this diagram that correspond to the dynamics observed at various times and places. [Reproduced from Earn et al [**ERBG00**]. Copyright © (2000) American Association for the Advancement of Science.]

Figure 12 presents the type of bifurcation diagram that we need, for the term-time forced SEIR model (Figure 9) with fixed parameters chosen to correspond to measles. The control parameter (on the abscissa) is the mean transmission rate $\langle \beta \rangle$. The ordinate shows measles incidence on 1 January of each year, so annual cycles are represented by a single curve, biennial cycles by double curves, and so on. Different colors correspond to different stable solutions of the model, which attract different sets of initial conditions (basins of attraction). For several values of the mean transmission rate $\langle \beta \rangle$, basins of attraction of the various attractors are shown above the bifurcation diagram in Figure 12. Where multiple stable solutions coexist, stochasticity can induce complicated dynamics due to shifts among attractors [**SS83, Sch85, Aro90**]. The upper panels of Figure 12 show that the basins of coexisting attractors are more intermixed if $\langle \beta \rangle$ is smaller, so we might expect the effects of stochasticity to be greater for smaller $\langle \beta \rangle$ (or, equivalently, when the effective $\langle \beta \rangle$ is reduced by vaccination or a decrease in birth rate).

Figure 13 shows measles incidence in four representative large cities; major dynamical transitions are evident in each case. Superimposed on the time series, each panel shows births (in red). In the vaccine era, susceptible recruitment (light blue) is lower than births. Above the incidence data, horizontal lines indicate periods of annual, biennial, and more complex dynamics that are predicted (Figure 12) on the basis of the observed exogenous variables (birth and vaccination rates). The horizontal lines in Figure 13 are color-coded according to the corresponding attractors in Figure 12.

London (Figure 13A) experienced biennial cycles of measles epidemics from 1950 to 1968; the estimated mean transmission rate for this period [**AM91**] is $\langle \beta \rangle \simeq 1240$, corresponding to a biennial attractor (dark blue, see Figure 12). Before 1950, epidemics were roughly annual; over the same brief period the birth rate was much higher, which greatly increased the effective mean transmission rate, allowing attraction to an annual cycle (green, far right of Figure 12). After 1968, recruitment rates steadily decreased because of mass vaccination (for example, when vaccine uptake reached 60%, the effective mean transmission rate was reduced to $\langle \beta \rangle \simeq 500$); this brought the system into the parameter region where there are multiple coexisting attractors with extremely intermixed basins, suggesting an explanation for the irregular epidemics in the vaccine era. Alternatively, or in addition, irregular dynamics in this region may arise from stochastic interactions with a chaotic repellor [**RW91**].

In Liverpool (Figure 13B) the birth rate was much higher than the mean in England and Wales throughout the post-war period until 1968 [**FG98**] (the birth rate in Liverpool is drawn as a dotted red line in Figure 13A as well, for comparison with the birth rate in London). This explains the roughly annual cycle of measles epidemics over the same period. After 1968, the combination of vaccination and a lower birth rate brought Liverpool, like London, into the regime where multiple co-existing cyclical attractors exist and might lead to irregular dynamics in the presence of noise.

Birth rates in the United States were relatively low during the Great Depression. Throughout this period, measles epidemics were irregular in New York and Baltimore (Figure 13C and D), consistent with stochastic switching between densely intermixed attractors or a chaotic repellor. After World War II, birth rates rose
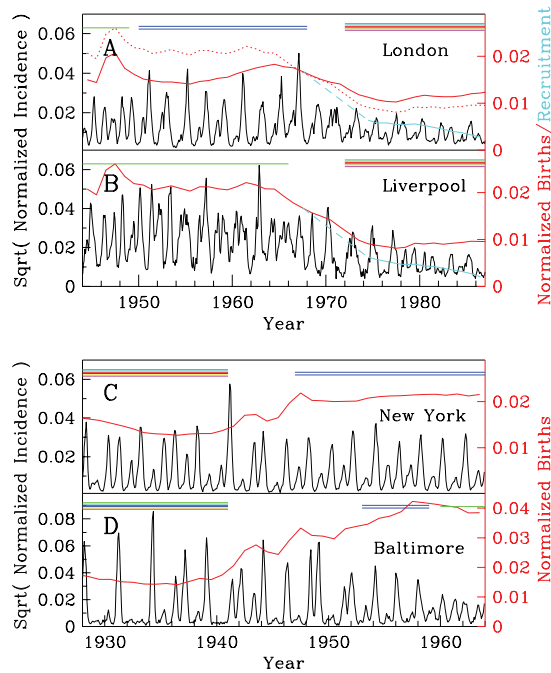
FIGURE 13. Measles dynamics in four large cities: London, Liverpool, New York, and Baltimore. In each panel, the black time series shows the square root of measles incidence normalized by the population size in 1960; the UK data are 4-weekly (13 points per year) whereas the US data are monthly (12 points per year). The red curve shows annual births normalized by the population size in 1960. In London and Liverpool, the light blue curve shows (normalized) annual susceptible recruitment (which is lower than the birth rate in the vaccine era). Together with the mean transmission rate $\langle\beta\rangle$, the recruitment rate determines the nature of the dynamics; see Eq. (4.2) (effects of changes in birth rate will be delayed by initial maternally acquired immunity and the age of school entry). The vertical scale for Baltimore is higher because reporting rates were twice as high [see London and Yorke [**LY73**]]. Note that the American data cover an earlier period than the English data; perhaps because of increases in mobility with time (as well as differences in population density), estimated values of $\langle\beta\rangle$ are significantly lower for the American cities [**And82, BE03**]. This (and the length of the summer holiday; see caption to Figure 12) must be borne in mind when comparing measles incidence data with various regions of Figure 12. The colored lines at the top of each panel correspond to the region of the bifurcation diagram in Figure 12 that applies to each section of the measles time series. Many colored lines together indicate the region of Figure 12 containing many coexisting attractors with intermixed basins. Two horizontal blue lines indicate the biennial attractor in Figure 12, while the single green line indicates the annual attractor that exists for high transmission rates. Gaps, where no region of Figure 12 is indicated, correspond to transient periods when the recruitment rate changed relatively rapidly. [Reproduced from Earn et al [**ERBG00**]. Copyright © (2000) American Association for the Advancement of Science.]

dramatically, pulling the system out of the regime with multiple co-existing attractors. In New York, the birth rate quickly reached a plateau, apparently fixing the system on the biennial attractor. In Baltimore, the birth rate continued to rise, eventually enough to bring the effective $\langle\beta\rangle$ into the region of Figure 12 where either biennial or annual cycles are possible (far right of Figure 12).

We now appear to be doing extremely well. We can successfully predict all of the dynamical transitions evident in each the measles time series in Figure 13 (and every other measles time series that I am familiar with). Note, incidentally, that this rules out chaotic intermittency as a mechanism for the dynamical transitions: if chaotic intermittency were the correct explanation for changes in periodicity of childhood disease epidemics then it would not be possible to predict the times at which changes in cycle structure occur.

Our confidence in the term-time forced SEIR model is now considerable, based on analysis of measles dynamics. Naturally, the next question we would like to address is whether the model is equally good at predicting the dynamics of other childhood diseases. We have every reason to anticipate success with other similar diseases, since the only relevant difference should be the values of the disease-specific parameters ($\mathcal{R}_0$, σ, γ).

Unfortunately, the analysis that has worked so well at explaining measles dynamics fails miserably for many other childhood infections!! In particular, with parameters estimated for rubella or whooping cough, the term-time forced SEIR model yields a unique attractor—an annual cycle—throughout the relevant range of $\langle\beta\rangle$. Yet the observed dynamics of these diseases is much more complicated than an annual cycle. Figure 14 shows examples of time series of case notifications for rubella and whooping cough, together with periodograms computed separately for two segments of each time series. It is clear from the time plots that the dynamics are not strictly annual for either disease. Moreover, there are strong spectral peaks at frequencies not predicted by the *asymptotic analysis*, i.e., not displayed by the (strictly annual) attractor of the term-time forced SEIR model.

Ugh! Could our success with measles have been extraordinarily lucky? Why doesn't the same approach lead to correct predictions of the dynamics of other childhood diseases?

Before giving up hope, it is quite telling that the transitions in the dynamics of rubella and whooping cough observed in Figure 14 correspond to substantial changes in the susceptible recruitment rate (plotted in green), which is exactly the parameter that we predicted would control dynamical transitions.

Once again, the truly discrete nature of disease transmission comes to our rescue. As Bartlett [**Bar57**] discovered for the unforced SIR model, demographic stochasticity can sustain oscillations that would be damped out in the absence of noise. In the seasonally forced model, noise can prevent the decay of transient oscillations at a period different from that of the attractor. We call the period associated with the attractor the *resonant period*, because it is arises from resonance with the seasonal forcing period of one year. In contrast, we say the period of transient oscillations is *non-resonant* because it is stimulated by demographic stochasticity and need not be an integer multiple of the period of seasonal forcing [**BE03**].

The resonant period is predicted by the asymptotic analysis that identifies the attractors of the model and worked so well for measles. The non-resonant period
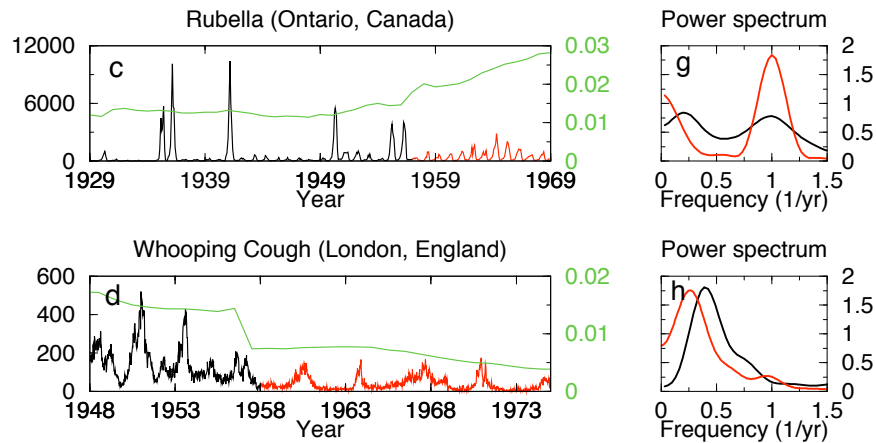
FIGURE 14. Case notification time series for rubella in Ontario, Canada, and whooping cough in London, UK. In each panel, the green line shows annual susceptible recruitment, $\nu(1-p)$, where ν denotes births normalized by 1955 population size and p denotes the proportion vaccinated; recruitment is shown displaced forward in time by the mean age at infection to account for the typical delay between birth and infection (11 years for rubella and 4 years for whooping cough). Time-series are divided into sections based on substantial differences in recruitment rates; the corresponding periodograms are not sensitive to the precise point at which the time-series is divided. [Reproduced from Bauch and Earn [**BE03**]. Copyright © (2003) The Royal Society.]

can be predicted by a *perturbation analysis*. Rather than a simple linear stability analysis of an equilibrium solution, we must consider perturbations of non-trivial attractors. Conducting such a perturbation analysis is not straightforward, but standard (numerical) methods exist for doing this [**Kuz98, BE03**].

Both the resonant and non-resonant periods can be expected to correspond to peaks in the periodogram of the disease time series. In every case I have ever considered, a resonant peak in the periodogram is correctly predicted by asymptotic analysis (and the prediction is always exact, which is perhaps not too surprising given that seasonal forcing will tend to lock the system onto an exact integer multiple of the forcing period). Moreover, a non-resonant peak in the periodogram is also predicted with great accuracy by perturbation analysis, for all of the childhood diseases we have investigated; see Figure 15 and Bauch and Earn [**BE03**].

We are now back in business. We have developed a quantitative method for predicting the structure of the periodogram of disease time series, and it works!

You may wonder why we didn't need to worry about sustaining transient dynamics in the case of measles. The answer is that the non-resonant period associated with the biennial attractor is too long to be observable in the relatively short time series we have. In this sense, we were indeed lucky in our analysis of measles dynamics using only the asymptotic analysis. On the other hand, restricting attention to bifurcation theory did lead us to what now appears to be an incorrect inference. At times when measles is not strictly biennial, we find that the periodogram has two peaks, a resonant peak at one year and non-resonant peak at the frequency
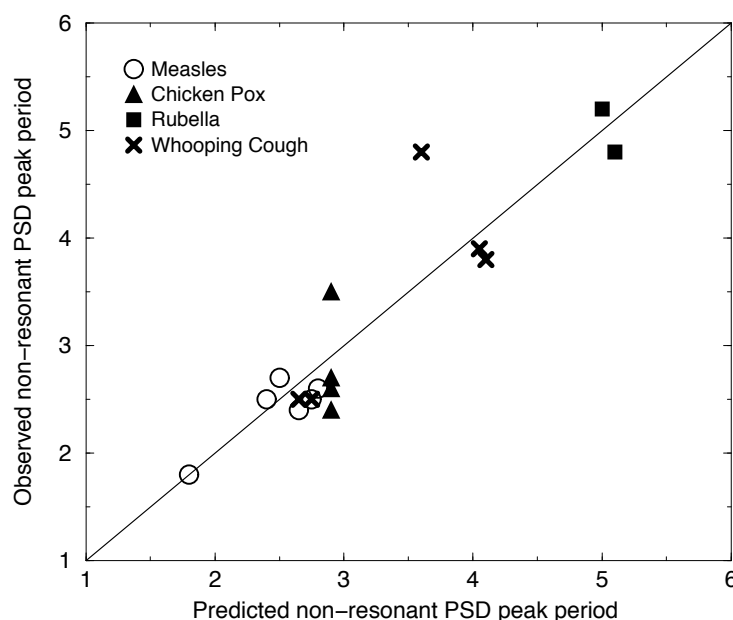
FIGURE 15. The correlation between predicted and observed periods of non-resonant periodogram peaks is positive and highly significant ($r^2 = 0.83$, $t = 8.69$, d.f. $= 16$, $p < 10^{-7}$). If the observations agreed perfectly with our predictions (as they do for resonant periodogram peaks) then the corresponding points would lie precisely on the line of slope 1 shown on the graph (the least-squares linear fit of the data has a slope of 1.02 and an intercept of 20.11). See electronic Appendix A of Bauch and Earn [**BE03**] for details. [Reproduced from Bauch and Earn [**BE03**]. Copyright © (2003) The Royal Society.]

associated with transient oscillations around the annual attractor. Apparent irregularity of measles epidemics does not, in fact, seem to arise from stochastic switching among co-existing attractors. We appear to be seeing simply the effects of noise preventing decay of transients.

While the approach we have adopted successfully predicts the positions of both peaks in the periodogram, it cannot explain the heights of the peaks. For that, we return again to demographic stochasticity. If the non-resonant peak occurs because noise sustains transient dynamics then that peak should be relatively larger if there is more noise. In other words, we expect that time series of infectious disease dynamics in smaller populations will show relatively larger non-resonant peaks, and this is what we observe, both in observed data and stochastic simulations [**BE03**].

4.1. Exercises

(1) Show, by scaling of the usual variables in the SEIR model as in Eq. (4.1), that a change in vaccination proportion from p to p' together with a change in birth rate from ν to ν' yields exactly the same dynamics as a change in basic reproduction number from $\mathcal{R}_0$ to $\mathcal{R}_0(1-p')\nu'/[(1-p)\nu]$, as indicated in Eq. (4.2).

(2) In the SIR model, show that if the birth rate ν is different from the death rate μ, and the contact term is $\beta SI/N$, rather than βSI, then the equations for the *proportions* of the population in each state are the same as the usual SIR model with both the birth rate and the death rate equal to ν. What if ν is a function of time? Is this consistent with the equivalence you derived in the previous exercise?

5. Manipulating epidemics

Vaccination can be viewed as a method of manipulating the structure of epidemic patterns. As a result of the equivalence (4.2) discussed in the previous section, vaccination at a given level will induce a predictable transition in the character of recurrent epidemics. Of course, if we vaccinate a sufficiently large proportion of the population (3.10) then we will eradicate the infection. But can we somehow build from our success in predicting dynamical transitions to develop a vaccination strategy that beats the classical eradication threshold (3.10)?

All of our analysis has been of temporal patterns of epidemics. What about spatial patterns? Can a spatially explicit version of the SEIR model lead us to suggest improved strategies for control or eradication of measles of other infectious diseases?

Before getting involved with a spatial model, it would make sense to look at some spatial epidemic data and see whether the data inspire any particular ideas. One potentially promising idea is suggested by a comparison of measles and whooping cough dynamics in England and Wales during the second half of the 20th century. Before mass vaccination for measles was initiated in 1968, measles epidemics in UK cities were strongly synchronized, but the degree of spatial correlation was greatly reduced in the vaccine era [**BG96**]. In sharp contrast, whooping cough epidemics in the identical population were not synchronized in the pre-vaccine era but became strongly synchronized upon implementation of mass vaccination [**REG99**]. These examples indicate that vaccination can induce spatial dynamical transitions (in the degree of synchony), in addition to the temporal transitions discussed in the previous section.

Attempting to manipulate the degree of synchrony of epidemics might be worthwhile. Increasing synchrony has the potential to reduce the threshold vaccination level required for eradication [**ERG98**], because synchrony reduces the "rescue effect" [**BKB77**] whereby regions in which populations are small are saved from extinction by dispersal of individuals from other areas. Can we devise a vaccination strategy that will synchronize measles epidemics worldwide?

One way to synchronize epidemics might be through "global pulse vaccination", i.e., intense vaccination efforts that are simultaneous around the world and repeated periodically [**ERG98**], or by supplementing continuous vaccination with periodic pulses [**GH97**]. Intuitively, the introduction of globally synchronized periodic forcing might synchronize the troughs between epidemics, stimulating the desired "anti-rescue effect". But this tantalizing prospect requires careful analysis. As we saw for the simple SIR model, periodic forcing can have complex dynamical effects, and we can't be sure *a priori* that the probability of synchronous oscillations will definitely be increased by global pulse vaccination.

The theory of synchronization is important in many areas of science [**PRK01, Str03**]. Most mathematical research on the subject has focused on the local stability of synchronous solutions of dynamical systems, which determines whether particular synchronous solutions attract nearly synchronous initial states with some non-zero probability [**Bue97**]. For epidemiological purposes, we really need a global theory, so we can determine conditions under which all initial states will lead to synchronous solutions [**ELR00**]. While we have made some valuable progress on a fully global theory of synchrony [**EL06**], it remains to be seen whether it can be successfully exploited to discover synchronizing vaccination strategies. Of course, a great deal might be learned from intermediate approaches that attempt to find conditions that yield synchrony with high probability but not certainty [**MEms**]. Simulations will also play an important role, though the parameter space that must be explored is high-dimensional and involves substantial computational challenges.

6. Conclusions and Take-home messages

The detailed temporal patterns of recurrent epidemics can be usefully described and analyzed using tools of time series analysis, especially the time plot and the periodogram. In terms of their ability to forecast patterns of recurrence of future epidemics, time series models such as $\text{ARMA}(p, q)$ models are of limited value. Mechanistic mathematical models, on the other hand, are extremely powerful in this respect. (Note that there is scope for combining time series models with mechanistic modelling, which leads to models that can fit observed epidemic time series with remarkable precision [**FG00**].)

Mathematical and computational analysis of a simple mechanistic model (the term-time forced SEIR model; Figure 9, Eqs. 3.9 and 3.17) explains the observed complex dynamics of epidemics of a variety of childhood diseases. We arrived at this model after ruling out a number of simpler models, and I hope that I have convinced you that successful mathematical modelling of biological systems is likely to proceed in this way. By adding model structure in small steps, we are most likely to discover which biological features have the greatest impact on observed behaviour. It is also worth highlighting that apparent limitations of a model can sometimes reflect more about our failure to think of the right analysis tool than about any inherent inadequacy of the model itself.

We are lucky that medical and public health personnel have painstakingly conducted surveillance of infectious diseases for centuries. This has created an enormous wealth of valuable data [**IIDDA**] with which to test hypotheses about disease spread using mathematical models. Building from models that successfully explain past patterns of epidemics, we can hope to design improved strategies for control or eradication of many infectious diseases.

Bibliography

[AB00] Andersson, H. and Britton, T., *Stochastic epidemic models and their statistical analysis*, Lecture notes in Statistics, vol. 151, Springer-Verlag, New York, 2000.

[AM91] Anderson, R. M. and May, R. M., *Infectious diseases of humans: Dynamics and control*, Oxford University Press, Oxford, 1991.

[And82] Anderson, R. M. (ed.), *The population dynamics of infectious diseases: Theory and applications*, Population and Community Ecology, Chapman and Hall, London, 1982.

[Aro90] Aron, J. L., *Multiple attractors in the response to a vaccination program*, Theoretical Population Biology **38** (1990), 58–67.

[Bai75]   Bailey, N. T. J., *The mathematical theory of infectious diseases and its applications*, Second ed., Hafner Press, New York, 1975.

[Bar57]   Bartlett, M. S., *Measles periodicity and community size*, Journal of the Royal Statistical Society, Series A **120** (1957), 48–70.

[Bar60]   ———, *Stochastic population models in ecology and epidemiology*, Methuen's Monographs on Applied Probability and Statistics, vol. 4, Spottiswoode, Ballantyne & Co. Ltd., London, 1960.

[BCC01]   Brauer, F. and Castillo-Chavez, C., *Mathematical models in population biology and epidemiology*, Texts in Applied Mathematics, vol. 40, Springer-Verlag, New York, 2001.

[BE03]   Bauch, C. T. and Earn, D. J. D., *Transients and attractors in epidemics*, Proceedings of the Royal Society of London, Series B **270** (2003), no. 1524, 1573–1578.

[Ber1760]   Bernoulli, D., *Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des advantages de l'inoculation pour la prévenir*, Mém Mathematical Physics Academy Royal Science Paris (1760), 1–45.

[BB04]   Bernoulli, D., *An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it*, Reviewed by Blower, S., Rev. Med. Virol. **14** (2004), 275-288, http://www.semel.ucla.edu/biomedicalmodeling/pdf/Bernoulli&Blower.pdf

[BG93]   Bolker, B. M. and Grenfell, B. T., *Chaos and biological complexity in measles dynamics*, Proceedings of the Royal Society of London, Series B **251** (1993), 75–81.

[BG96]   ———, *Impact of vaccination on the spatial correlation and persistence of measles dynamics*, Proceedings of the National Academy of Sciences, USA **93** (1996), 12648–12653.

[BKB77]   Brown, J. H. and Kodric-Brown, A., *Turnover rates in insular biogeography: effect of immigration on extinction*, Ecology **58** (1977), 445–449.

[Bue97]   Buescu, J., *Exotic attractors: From Liapunov stability to riddled basins*, Progress in Mathematics, vol. 153, Birkhäuser Verlag, Basel, 1997.

[Cha04]   Chatfield C., *The analysis of time series: An introduction*, sixth ed., Texts in Statisitical Science Series, Chapman & Hall/CRC, Boca Raton, 2004.

[DPV+06]   Dushoff, J., Plotkin, J. B., Viboud, C., Earn, D. J. D., and Simonsen, L., *Mortality due to influenza in the United States - an annualized regression approach using multiple-cause mortality data*, American Journal of Epidemiology **163** (2006), no. 2, 181–187.

[EDL02]   Earn, D. J. D., Dushoff, J., and Levin, S. A., *Ecology and evolution of the flu*, Trends in Ecology and Evolution **17** (2002), no. 7, 334–340.

[EL06]   Earn, D. J. D. and Levin, S. A., *Global asymptotic coherence in discrete dynamical systems*, Proc. Natl. Acad. Sci. USA **103** (2006), no. 11, 3968–3971, [*Correction:* PNAS **103** (2006), no. 51, 19605].

[ELR00]   Earn, D. J. D., Levin, S. A., and Rohani, P., *Coherence and conservation*, Science **290** (2000), no. 5495, 1360–1364.

[ERBG00]   Earn, D. J. D., Rohani, P., Bolker, B. M., and Grenfell, B. T., *A simple model for complex dynamical transitions in epidemics*, Science **287** (2000), no. 5453, 667–670.

[ERG98]   Earn, D. J. D., Rohani, P., and Grenfell, B. T., *Persistence, chaos and synchrony in ecology and epidemiology*, Proceedings of the Royal Society of London, Series B **265** (1998), no. 1390, 7–10.

[FBBC97]   Fitch, W. M., Bush, R. M., Bender, C. A., and Cox, N. J., *Long term trends in the evolution of H(3) HA1 human influenza type A*, Proceedings of the National Academy of Sciences, USA **94** (1997), no. 15, 7712–7718.

[FG98]   Finkenstädt, B. F. and Grenfell, B. T., *Empirical determinants of measles metapopulation dynamics in england and wales*, Proceedings of the Royal Society of London, Series B **265** (1998), 211–220.

[FG00]   Finkenstädt, B. F. and Grenfell, B. T., *Time series modelling of childhood diseases: A dynamical systems approach*, Applied Statistics **49** (2000), no. 2, 187–205.

[GBK01]   Grenfell, B. T., Bjornstad, O. N., and Kappey, J., *Travelling waves and spatial hierarchies in measles epidemics*, Nature **414** (2001), no. 6865, 716–723.

[GH83]   Guckenheimer, J. and Holmes, P., *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, Applied Mathematical Sciences, vol. 42, Springer-Verlag, New York, 1983.

[GH97]   Grenfell, B. and Harwood, J., *(meta)-population dynamics of infectious diseases*, Trends in Ecology and Evolution **12** (1997), 395–399.

[Gil76]   Gillespie, D. T., *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, Journal of Computational Physics **22** (1976), 403–434.

[Gle87]   Gleick, J., *Chaos*, Abacus, London, 1987.

[HHGL$^+$06] Hallett, D. H., Gleason, A. M., Lock, P. F., Flath, D. E., Lomen, D. O., Lovelock, D., McCallum, W. G., Osgood, B. G., Quinney, D., Rhea, K., Tecosky-Feldman, J., and Tucker, T. W., *Applied calculus*, third ed., John Wiley & Sons, New Jersey, 2006.

[IIDDA]   IIDDA, *The International Infectious Disease Data Archive*, http://iidda.mcmaster.ca.

[KM27]   Kermack, W. O. and McKendrick, A. G., *A contribution to the mathematical theory of epidemics*, Proceedings of the Royal Society of London, Series A **115** (1927), 700–721.

[KM04]   Korobeinikov, A. and Maini, P. K., *A Lyapunov function and global properties for SIR and SEIR epidemiological models with nonlinear incidence*, Mathematical Biosciences and Engineering **1** (2004), no. 1, 57–60.

[Kur80]   Kurtz, T. G., *Relationships between stochastic and deterministic population models*, Lecture Notes in Biomathematics **38** (1980), 449–467.

[Kuz98]   Kuznetsov, Y. A., *Elements of applied bifurcation theory*, Second ed., Applied Mathematical Sciences, vol. 112, Springer-Verlag, New York, 1998.

[LM95]   Li, M. Y. and Muldowney, J. S., *Global stability for the SEIR model in epidemiology*, Mathematical Biosciences **125** (1995), no. 2, 155–164.

[LY73]   London, W. and Yorke, J. A., *Recurrent outbreaks of measles, chickenpox and mumps. I. Seasonal variation in contact rates*, American Journal of Epidemiology **98** (1973), no. 6, 453–468.

[MEms]   McCluskey, C. and Earn, D. J. D., *Attractivity of coherent manifolds in metapopulation models*, (in prep).

[MM04]   Moote, A. L. and Moote, D. C., *The Great Plague: The story of London's most deadly year*, John Hopkins University Press, Baltimore, 2004.

[Nie01]   Nievergelt, Y., *Wavelets made easy*, Birkhäuser, Boston, 2001.

[NM00]   Nowak, M. A. and May, R. M., *Virus dynamics: Mathematical principles of immunology and virology*, Oxford University Press, Oxford, 2000.

[OS90]   Olsen, L. F. and Schaffer, W. M., *Chaos versus noisy periodicity: alternative hypotheses for childhood epidemics*, Science **249** (1990), 499–504.

[PRK01]   Pikovsky, A., Rosenblum, M., and Kurths, J. (eds.), *Synchronization: A universal concept in nonlinear sciences*, Cambridge Nonlinear Science Series, vol. 12, Cambridge University Press, Cambridge, 2001.

[PST93]   Platt, N., Spiegel, E. A., and Tresser, C., *On-off intermittency: a mechanism for bursting*, Physical Review Letters **70** (1993), no. 3, 279–282.

[PTVF92]   Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., *Numerical recipes in C: The art of scientific computing*, 2nd ed., Cambridge University Press, New York, 1992.

[REG99]   Rohani, P., Earn, D. J. D., and Grenfell, B. T., *Opposite patterns of synchrony in sympatric disease metapopulations*, Science **286** (1999), no. 5441, 968–971.

[Ros11]   Ross, B., *The prevention of malaria*, 2nd ed., Murray, London, 1911.

[RW91]   Rand, D. A. and Wilson, H. B., *Chaotic stochasticity: a ubiquitous source of unpredictability in epidemics*, Proceedings of the Royal Society of London, Series B **246** (1991), 179–184.

[Sch84]   Schenzle, D., *An age-structured model of pre- and post-vaccination measles transmission*, IMA Journal of Mathematics Applied in Medicine and Biology **1** (1984), 169–191.

[Sch85]   Schwartz, I. B., *Multiple stable recurrent outbreaks and predictability in seasonally forced nonlinear epidemics*, Journal of Mathematical Biology **21** (1985), 347–361.

[Scu03]   Scuffham, P. A., *Estimating influenza-related hospital admissions in children and adults – a time series analysis*, Disease Management and Health Outcomes **11** (2003), no. 4, 259–269.

[SK85]   Schaffer, W. M. and Kot, M., *Nearly one dimensional dynamics in an epidemic*, Journal of Theoretical Biology **112** (1985), 403–427.

[SKTO93]   Schaffer, W. M., Kendall, B. E., Tidd, C. W., and Olsen, L. F., *Transient periodicity and episodic predictability in biological dynamics*, IMA Journal of Mathematics Applied in Medicine and Biology **10** (1993), 227–247.

[SS83]   Schwartz, I. B. and Smith, H. L., *Infinite subharmonic bifurcation in an SEIR model*, Journal of Mathematical Biology **18** (1983), 233–253.

[Str94]   Strogatz, S. H., *Nonlinear dynamics and chaos*, Addison Wesley, New York, 1994.

[Str03]   _____, *Sync: The emerging science of spontaneous order*, Hyperion, New York, 2003.

[TC98]   Torrence, C. and Compo, G. P., *A practical guide to wavelet analysis*, Bulletin of American Meteorological Society **79** (1998), no. 1, 61–78. http://atoc.colorado.edu/research/wavelets/.

[USVS]   United States Bureau of the Census, *Vital statistics of the united states (before 1937: Mortality statistics)*, 1900-1992.

[Wig03]   Wiggins, S., *Introduction to applied nonlinear dynamical systems and chaos*, 2nd ed., Texts in applied mathematics, vol. 2, Springer-Verlag, New York, 2003.