Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 223 (2003) 19–25
Journal of Theoretical Biology
www.elsevier.com/locate/jtbi

Interpreting time-series analyses for continuous-time biological models—measles as a case study

K. Glass*, Y. Xia, B.T. Grenfell

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

Received 12 June 2002; received in revised form 27 November 2002; accepted 4 December 2002

Abstract

An increasing number of recent studies involve the fitting of mechanistic models to ecological time-series. In some cases, it is necessary for these models to be discrete-time approximations of continuous-time processes. We test the validity of discretization in the case of measles, where time-series models have recently been developed to estimate ecological parameters directly from data. We find that a non-homogeneous contact function is necessary to capture the host–parasite interaction in a discrete-time model, even in the absence of heterogeneities due to spatial or age structure. We derive a mathematical relationship describing the expected departure from mass-action transmission in terms of the epidemiological parameters in the model, and identify conditions under which the discretization process may fail.

Keywords: Epidemiological models; Measles; Time-series analysis; Population dynamics

1. Introduction

Continuous-time models, both deterministic (Anderson and May, 1979, 1991; Murray, 1989) and stochastic (Bartlett, 1960, 1961; Mollison et al., 1994) have a long history in the study of biological systems. Relatively simple sets of coupled equations may be used to model the mechanisms underlying a continuously changing system. In many real-world systems, it is difficult to measure (or even identify) all state variables, and instead measurements are made of a single variable at discrete time intervals. This time-series can then be fitted to a mechanistic discrete-time model, of which the Nicholson–Bailey host–parasite model (Nicholson and Bailey, 1935) is an early example. While discrete-time models are often better suited to the available data, not all simple discrete-time models successfully reproduce the dynamics of the standard continuous-time models (Mollison and Din, 1993). It is important to examine the effect that the discretization process has on model behaviour before conclusions can be drawn from the time-series analysis.

To address this question, we need to analyse systems for which rich time-series are available, and which are sufficiently well understood for realistic discrete- and continuous-time models to be made. Measles is a perfect case study for this investigation, with its extensive notification time-series (Grenfell and Harwood, 1997) and relatively simple natural history. A family of continuous-time models based on the susceptible–infective–recovered (SIR) paradigm successfully capture the recurrent pre-vaccination sequence of measles epidemics and the impact of vaccination (Anderson and May, 1991; Schenzle, 1984; Earn et al., 2000). Early attempts to construct discrete-time models of measles (Fine and Clarkson, 1982) have been found to display unsatisfactory dynamics (Mollison and Din, 1993), failing to capture the strong biennial cycle seen in numerous pre-vaccination data sets (e.g. England and Wales, New York, Baltimore). A recently developed discrete-time model allows for the estimation of epidemiological parameters (notably the seasonality in transmission rate) directly from disease time-series (Finkenstädt and Grenfell, 2000). This time-series SIR

*Corresponding author. National Centre for Epidemiology and Population Health, Australian National University, Canberra ACT 0200, Australia. Tel.: +61-2-6125-2468.

E-mail address: katie@zoo.cam.ac.uk, kathryn.glass@anu.edu.au (K. Glass).

0022-5193/03/$ - see front matter
doi:10.1016/S0022-5193(03)00031-6

(TSIR) model successfully bridges the gap between mechanistic models and time-series data, capturing many of the salient features of pre-vaccination measles in large cities (Bjørnstad et al., 2002; Grenfell et al., 2002). In the light of the earlier difficulties with discrete-time models, a thorough investigation of the dynamical behaviour of the TSIR model over the range of parameter values seen for measles is crucial for interpreting the time-series analysis.

Of particular interest are the intriguing departures from homogeneous 'mass-action' transmission discovered in Finkenstädt and Grenfell (2000). These findings could represent important evidence of population manifestations of spatial heterogeneity in infection. However, we first need to examine whether the discretization of the dynamics inherent in this time-series model affects estimates of mass-action transmission.

This paper begins with a discussion of current measles models, and then provides a comprehensive comparison of the time-series model with the more established continuous-time model. It becomes apparent that the use of a non-homogeneous contact function in discrete time is crucial for modelling mass-action transmission in continuous time. We provide a mathematical analysis of the relationship between the contact functions of the two models, and consider the consequences of significant changes to the system such as that induced by measles vaccination. Our findings raise important issues for a large class of discrete-time ecological models.

2. Models

During the pre-vaccination era, measles dynamics in large cities were predominantly periodic, showing annual, triennial, but most commonly biennial cycles (Grenfell et al., 2002). Following the introduction of mass vaccination in the 1960s, measles dynamics became more irregular, and possibly chaotic (Godfray and Grenfell, 1993). These varied dynamics can be captured by a very simple continuous-time SIR model (see Anderson and May, 1991; Schenzle, 1984):

$$\begin{aligned}\dot{S} &= -b(t)S(t)I(t) + m(1 - S(t)),\\ \dot{I} &= b(t)S(t)I(t) - (m+g)I(t),\\ \dot{R} &= gI(t) - mR(t),\end{aligned} \tag{1}$$

where m is the birth rate, g the recovery rate and $b(t)$ is a time-varying transmission parameter of the form $b(t) = b_0(1 + b_1\cos(2\pi t))$ with period 1 year, designed to model the variability in transmission rate induced by the school year. In the above form, S, I and R represent the fraction of the population that is susceptible, infected and recovered at any time, and the population size is assumed constant (thus $S + I + R = 1$) allowing us to concentrate only on the susceptible and infected dynamics. The model is simulated with $g = 0.077$, $b_0 = 1.308$ and $b_1 = 0.08$, corresponding to the epidemiological parameters for measles of an exposed and infectious period of 13 days, a reproductive ratio of 17, and 8% seasonal forcing.

Earn et al. (2000) demonstrated that both the between-city variability in period of oscillation before vaccination, and the within-city transition following vaccination, can be explained as bifurcations of this nonlinear model in a single parameter. The bifurcation diagram in Fig. 1(a) shows the stable periodic cycles for the SIR model for m in the range 0–0.05. Before vaccination, cities with high birth rates (e.g. Liverpool) showed annual cycles and cities with lower birth rates (e.g. London) showed biennial cycles. Following vaccination (which effectively drops the susceptible recruitment rate), both cities display irregular dynamics that can be explained as switching between multiple stable cycles (see Earn et al. (2000); Keeling et al. (2001) for more detail on this).

While simple continuous-time models have been highly successful in reproducing the dynamical behaviour seen in measles time-series, they are not particularly suitable if the parameters of sub-models of biological processes (such as seasonality in transmission and effects of spatial and age structure in the population) are to be estimated from time-series data. In order to better exploit the available data, Finkenstädt and

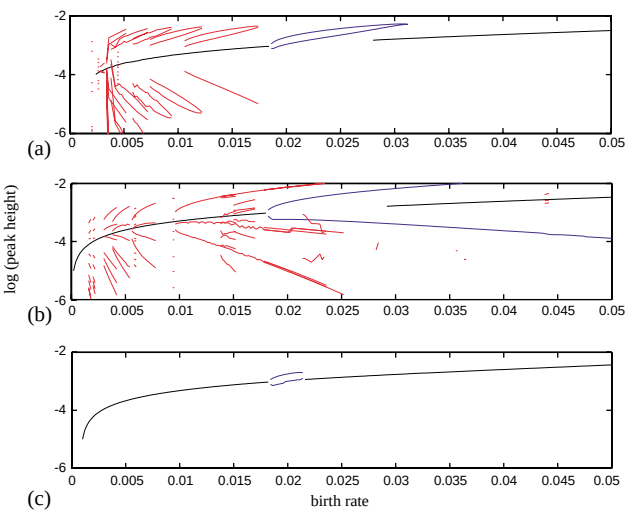

Fig. 1. Bifurcation diagrams of yearly peak height of the proportion of infected individuals against birth rate for (a) the continuous-time SIR model (1), (b) the discrete-time model (3) with $A = 1$, and (c) the discrete-time model with $A = 0.976$. Annual cycles are shown in black, biennial cycles in blue and higher period cycles in red. The cycles shown are the stable orbits found after numerical simulations of the model for a range of initial conditions.

Grenfell (2000) introduced the stochastic discrete-time TSIR model:

$$S_t = M_{t-d} + S_{t-1} - I_t + u_t,$$
$$I_t = \beta_t I_{t-1}^{\alpha} S_{t-1}^{\gamma} \varepsilon_t, \quad (2)$$

where I_t and S_t are the numbers of infectious and susceptible individuals, respectively, at time t, β_t is a time-varying transmission parameter with period 1 year, M_{t-d} denotes births (time-delayed to take account of maternal immunity), α and γ are scalars estimated from data, and u_t, ε_t are random variables with mean zero. For data from England and Wales, the parameter α was estimated to be 0.97 (slightly but significantly less than 1), and this model has subsequently been used to analyse spatio-temporal waves of infection in England and Wales (Grenfell et al., 2001).

To compare the stochastic TSIR model (2) with the deterministic continuous-time SIR model (1), we ignore the stochastic components (u_t, ε_t) and use the deterministic skeleton:

$$S_{t+T} = S_t - I_{t+T} + M(1 - S_t),$$
$$I_{t+T} = B(t) S_t I_t^A. \quad (3)$$

Here, S_t and I_t now represent the fraction of the population that is susceptible and infected, respectively, at time t, T is a time step of 2 weeks dictated by the sampling period of the data, and A is a scalar ($A = 1$ giving the standard mass-action transmission). Individuals are infected for a single time step of 14 days, which matches the epidemiological assessment of 13 days reasonably well. As Finkenstädt and Grenfell (2000) were not able to estimate the value of the exponent γ in (2), and the model is not sensitive to changes in this parameter, we assume $\gamma = 1$. To remove any effects of changes in the birth rate over time, we use a fixed birth rate, M (maintaining a constant population size). $B(t)$ is a time-varying function of the form $B(t) = B_0(1 + B_1 \cos(2\pi t))$ designed to model the seasonal variation in transmission induced by the school term. Representing the annual variance in transmission rates as a sinusoidal function is a simplification of the underlying real-world dynamics. However, while on–off functions that mimic the school term (Schenzle, 1984) fit the England and Wales data more accurately, when comparing models (1) and (3), modifying the form of the function did not change the results. All three models described here combine exposed (i.e. infected but not yet infectious) and infectious individuals in one class. Although susceptible–exposed-infectious–recovered (SEIR) models are more biologically realistic, their dynamical behaviour does not differ significantly from SIR models in the case of measles (Keeling et al., 2001).

3. Results

For a birth rate of $m = 0.02$ (corresponding to $M = 0.02T$) the continuous-time model (1) exhibits biennial cycles, as do the observed time-series in Earn et al. (2000). However, the discrete-time model (3) under the mass-action assumption (that is, with $A = 1$), exhibits different periodic cycles for different initial conditions (see Fig. 1(a) and (b) with $m = 0.02$). If the mass-action assumption is relaxed, and A is varied to fit the discrete-time biennial cycle to that of the continuous-time model using least squares (with all other parameters fixed), A is estimated to be 0.976, and the discrete-time model displays biennial cycles only (see Fig. 1(c) with $m = 0.02$). For these parameter values at least, a non-homogeneous contact function in discrete time better mimics homogeneous mass-action in continuous time. It might be supposed that some of the discrepancy between the discrete- and continuous- time SIR models lies in the distribution of the infectious period, which is exponential in the case of model (1) and has a fixed length in the case of (3). Re-estimating A for a continuous-time SEIR model with fixed infectious period (see Keeling and Grenfell (1997) for details of this model) gave an estimate of 0.982. For comparison, the estimation of A for the SEIR model with exponentially distributed infectious period was 0.976. The discrete-time model with each of these values of A has a unique stable (biennial) cycle.

Analysis of the discrete-time model with constant B indicates that reducing A from 1 increases the stability of the equilibrium. This leads to a greater tolerance to environmental noise. Fig. 2 shows the perturbation to

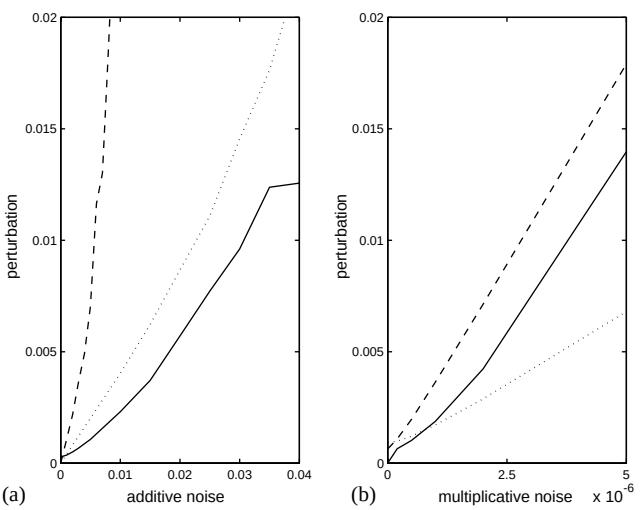

Fig. 2. The perturbation to the biennial cycles of the continuous-time model (—), the discrete-time model with $A = 1$ (- - -) and the discrete-time model with $A = 0.976$ (···) under (a) additive and (b) multiplicative noise. The perturbation is measured as the sum of squares of the difference in the number of infectives of the noisy and noise-free cycles, averaged over 100 simulations.

the biennial cycle produced by introducing either multiplicative or additive noise into the I_t component of the model (thus either $I_t(1+\mu_t)$ or $I_t+\mu_t$, where μ_t has mean zero). Clearly, the discrete-time model has a higher tolerance to noise with $A<1$.

This increased stability has a dramatic effect on the bifurcation diagram (Fig. 1(c)). Although the discrete-time model with $A=0.976$ has a qualitatively similar structure to the continuous-time model in the neighbourhood of $m=0.02$, it does not have the same complicated bifurcation structure seen for low birth rates. On the other hand, the discrete-time model with mass-action transmission *does* exhibit the multiple co-existing cycles for low birth rates. An additional unexpected property of the discrete-time model with fixed $A<1$ is that the mean proportion of susceptibles varies with birth rate to a much greater degree than that of the continuous-time SIR model, the continuous-time SEIR model or the continuous-time SEIR model with fixed infectious period.

An explanation for these results is found by deriving a Taylor-series expansion for the dynamics of the continuous-time SIR model about the equilibrium, and equating terms in the expansion with those of an expansion of the discrete-time model, as described in detail in the appendix. A first-order expansion in the time step returns the mass-action solution. When second-order terms are included, however, the mass-action assumption breaks down. Equating the largest terms in the expansion produces the relationship

$$A = 1 - \tfrac{1}{2}mT^2(b-m-g).$$

If the time step T is approximately equal to the infectious period, this relationship can be expressed as

$$A \approx 1 - \tfrac{1}{2}M(R_0-1)$$
$$\approx 1 - \tfrac{1}{2}*(\textit{recruitment rate})*(\textit{interaction rate}),$$

where R_0 is the reproductive ratio of the disease. Thus, we expect the transmission term to differ most from mass-action when there is a high birth-rate, and approach mass-action when recruitment rates are low, as seen following vaccination.

A more accurate solution of the Taylor-series expansion can be found by equating all second-order terms in the expansion. In this case, there is no simple algebraic solution, but numerical solutions show the same relationship between A and birth rate. Fig. 3(a) gives the numerical solutions for A for a range of birth rates, m (with b, g and T fixed) and Fig. 3(b) shows the values of A for fixed m, g and T, letting $b(t)$ vary sinusoidally over 1 year. We see that A decreases with birth rate, and varies over the year proportionally with $-b(t)$. If parameter values are assigned according to this method, the bifurcation structure and mean level of susceptibles

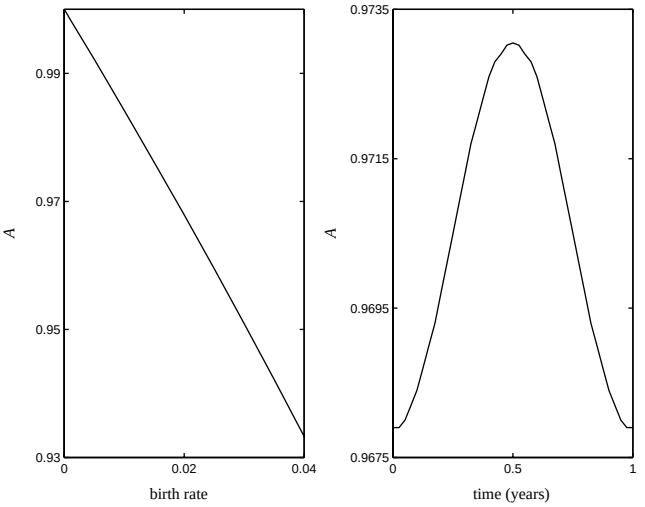

Fig. 3. Optimal values of A calculated using a Taylor-series approximation (a) with varying birth-rate and (b) over an annual cycle with sinusoidally forced transmission rate.

reproduces that of the continuous-time model much more closely.

While this result is specific to a comparison of the discrete-time model with simple continuous-time models, a search through phase space for regions of biennial cycles and higher order cycles demonstrates that as A decreases from 1 the model passes from higher order cycles to biennial cycles to annual cycles (see Fig. 4). Increasing the level of seasonal forcing expands the region of interest somewhat, but maintains the same structure.

In Fig. 5, we apply these results to data from London. Fig. 5(a) shows the observed bi-weekly cases corrected for under-reporting (see Bjørnstad et al. (2002) for details of the data). Superimposed on this time-series is the effective birth rate for London (that is, the birth rate adjusted by the vaccination rate) scaled to fit the y-axis. Applying the logarithm transformation (Box and Cox, 1964) the TSIR model (2) becomes a linear regression model, from which we estimate the value of α for the pre-vaccination London data to be 0.97. Fig. 5(b) and (d) show the deterministic realization and bifurcation diagram for the TSIR model with $\alpha=0.97$. We see that the model captures the annual–biennial transition with the change in birth rate in the pre-vaccination data, but shows stable annual cycles post-vaccination. For comparison, the TSIR model is simulated with a higher α (Fig. 5(c) and (e)). In the pre-vaccination era, the amplitude of oscillation is too high, and the fit to the data does not mimic the dynamics of the real system as well as that of $\alpha=0.97$. For the post-vaccination era, however, the higher value of α leads to dynamics that switch between multiple orbits, which shows a better qualitative agreement with the data.

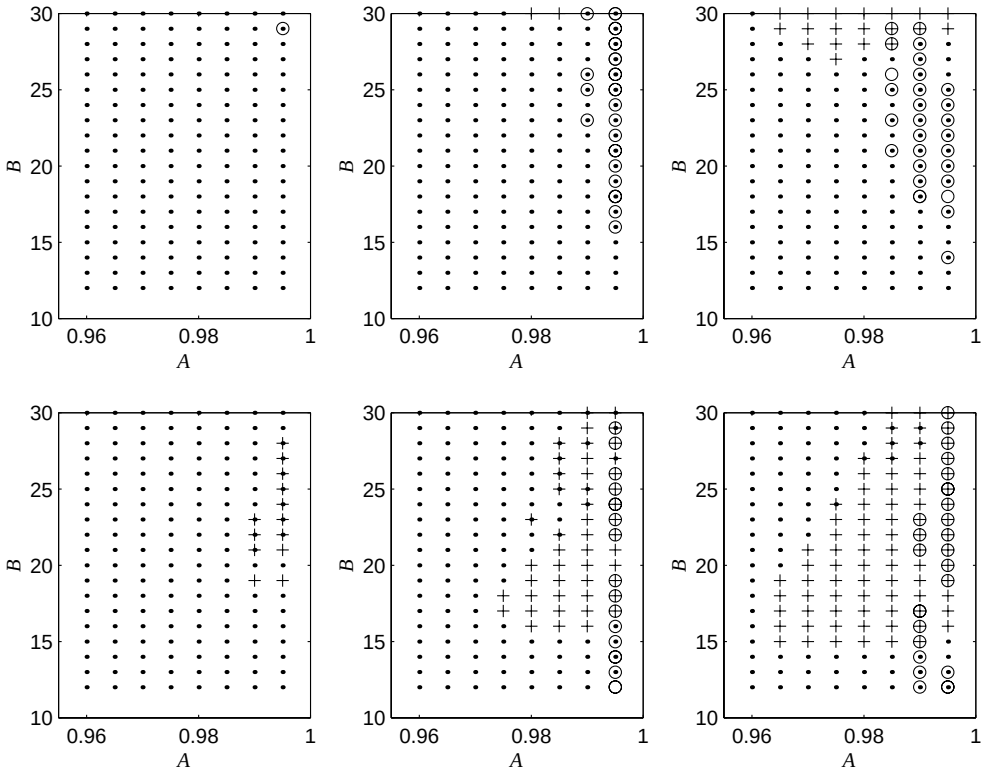

Fig. 4. Plots showing the distribution of periodic cycles in parameter space. Top row shows birth rate of 0.01, bottom row shows birth rate of 0.02. Columns from left to right have 4%, 8% and 12% seasonal forcing. Cycles are shown as: annual (·), biennial (+) and cycles of period 3 and above (◦).

4. Discussion

This study highlights a modelling difficulty that is relevant to a large class of biological systems. While real-world interactions occur continuously in time, data are sampled only at fixed time points. Discrete-time models that are well suited to estimating biological parameters from such time-series may fail to capture the continuous-time dynamics fully. In the case of measles, where simple continuous-time models reproduce very successfully many facets of the epidemiological dynamics, it is possible to consider in detail the consequences of the discretization process.

Firstly, as has already been noted by Mollison and Din (1993), the simplest discretization fails to exhibit the appropriate dynamics. In order to successfully reproduce pre-vaccination measles dynamics in discrete time, it is necessary to collapse the mass-action assumption. An explanation for this result may lie in the slow buildup of infection in a discrete-time model with mass-action. At the start of an epidemic, when numbers of infected are low and numbers of susceptibles high, the continuous-time model displays a sharp increase in the number of infected individuals. With a 2-week time step, the discrete-time model cannot display this form of rapid feedback, and under mass-action, infection builds up more slowly. The non-homogeneous transmission function gives increased weight to small numbers of infectives, thus allowing an epidemic to take off more rapidly. If the time step was to be increased further (from fortnightly to monthly, for example), the analysis suggests that there would be an even greater departure from mass-action transmission. While the exact relationship between the time step and the transmission function is specific to the measles model, this result highlights an issue that is generic to models of predator–prey interactions (the Lotka–Volterra model, for example). Simple discretizations of continuous-time models will work well if the time step remains short. As the difference between the time step and the response time of the contact function increases, however, it may become necessary to modify the contact function to model the predator–prey interactions accurately.

A caveat that arises from this study is that while spatial heterogeneity and age structure can result in contact rates that do not follow a simple mass-action form, the fact that a non-homogeneous transmission function is required to describe a time-series does not necessarily imply that there are heterogeneities in the system. In the case of pre-vaccination measles, we have demonstrated that a non-homogeneous discrete-time model is required to reproduce dynamics of a homogeneous continuous-time model. A value of $\alpha<1$ estimated from time-series data does not necessarily

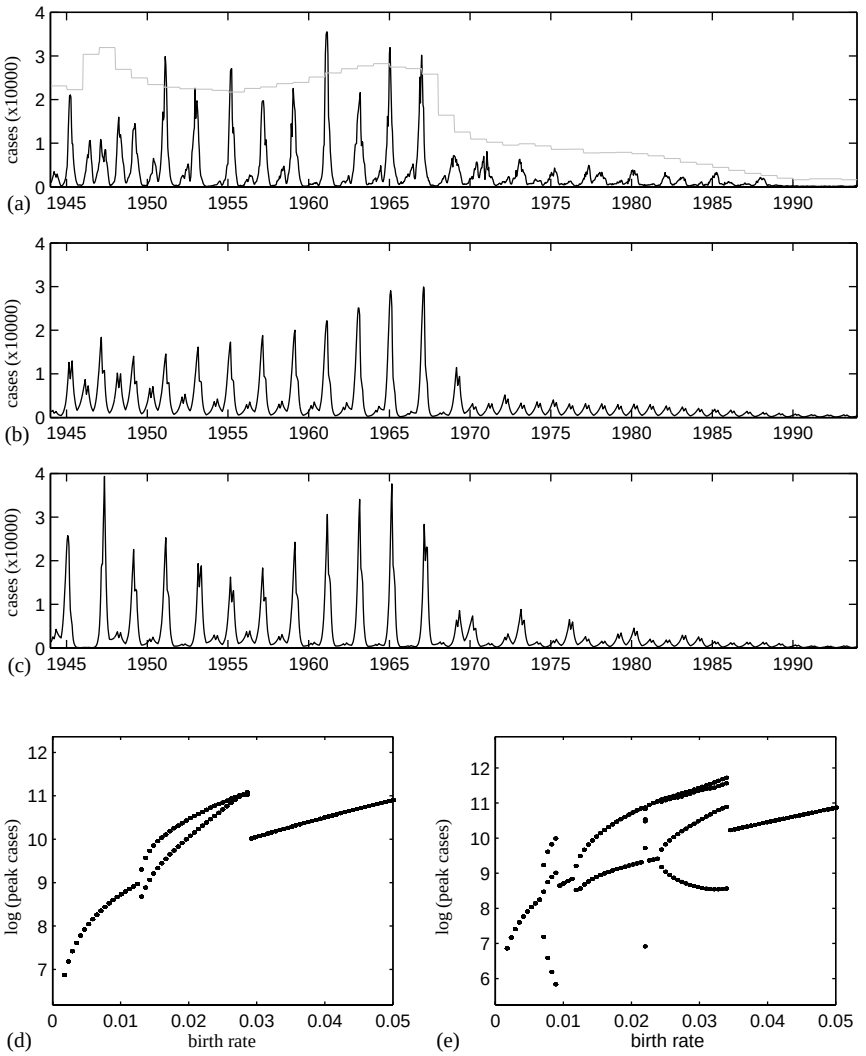

Fig. 5. A comparison of the TSIR model with measles data from London. (a) shows the observed bi-weekly cases, with the effective birth rate (the birth rate adjusted by vaccinations) superimposed above it. (b) and (d) show the time-series and bifurcation diagram, respectively, for the TSIR model with $\alpha = 0.97$. (c) and (e) give the time-series and bifurcation diagram for the model with $\alpha = 0.99$.

require a mechanistic interpretation, as it may arise as an artifact of the discretization process.

A further lesson is that models must capture the appropriate biology. Here, models that give very similar results when compared for one part of parameter space are dramatically different in others. It is clear that for greatly altered birth-rate, or data collected after vaccination, it is necessary to re-estimate the contact function. By finding an approximation to the contact function required to mimic a simple homogeneous model with no age structure, we provide a baseline function that can be compared against those estimated from real-world data. Although the relationship found here is specific to the SIR model, we have seen that the post-vaccination measles data are more closely reproduced by dynamics that switch between cycles of different periods than by stable annual or biennial cycles. The general structure of orbits seen in Fig. 4 suggests that when the TSIR model is used to simulate post-vaccination measles dynamics, either the transmission function should be adjusted to be closer to mass-action or the amplitude of seasonal forcing should be increased.

Acknowledgements

The authors would like to thank Pej Rohani, Matt Keeling and Bärbel Finkenstädt for useful discussions, and two anonymous referees for helpful comments. Financial support was provided by the Wellcome Trust.

Appendix

To derive the Taylor-series expansion, we assume $I_t = \bar{I} + J_t$, $S_t = \bar{S} + K_t$, $I(t) = \hat{I} + J(t)$, $S(t) = \hat{S} + K(t)$,

where $(\bar{S},\bar{I})$ and $(\hat{S},\hat{I})$ are the equilibria of the TSIR and the continuous-time SIR models, respectively. Note that unless $T=1/(m+g)$, these equilibria may not coincide exactly. We instead consider the dynamics about the equilibria by finding expansions for J_{t+T} and $J(t+T)$ of the form

$$J_{t+T}=C_1J_t+C_2K_t+C_3J_tK_t+C_4J_t^2+C_5K_t^2,$$

$$J(t+T)=c_1J(t)+c_2K(t)+c_3J(t)K(t)+c_4J(t)^2+c_5K(t)^2$$

using a Taylor-series expansion in the case of the continuous-time model, and a binomial expansion in the case of the discrete-time model. In the above equations, $c_i(b,m,g,T)$ and $C_i(B,A,M,T)$ are series in T and we assume that $M=mT$. Considering only first-order terms in T we have

$$c_1=1,\quad c_2=mT\left(\frac{b}{m+g}-1\right),$$
$$c_3=bT,\quad c_4=c_5=0,$$

$$C_1=A,\quad C_2=B\bar{I}^A,\quad C_3=\frac{A}{\bar{S}},$$
$$C_4=\frac{A(A-1)}{2\bar{I}},\quad C_5=0.$$

Equating $c_i=C_i$ we obtain $A=1$, $B(t)=b(t)T$, with an exact fit if $T=1/(m+g)$. When second-order terms (i.e. those involving T^2) are included, there is no simple algebraic solution for A and B if all five terms are used. Equating $c_1=C_1$ and $c_2=C_2$ only, we derive

$$A=1-\tfrac{1}{2}mT^2(b-m-g),$$

$$B=\frac{bT(1+\frac{1}{2}mT-mT\frac{b}{m+g})}{1-\frac{1}{2}mT^2(b-m-g)}.$$

Note that A is a decreasing function of m. Finding the optimal fit to all five terms can be found numerically by minimizing $\sum_{i=1}^{5}(c_i-C_i)^2$ over A and B for fixed values of b, g, m and T, assuming $M=mT$. Fig. 3(a) gives the optimal values of A for a range of birth rates, m (with b, g and T fixed) and Fig. 3(b) shows the optimal A for fixed m, g and T, letting $b(t)$ vary sinusoidally over 1 year.

References

Anderson, R.M., May, R.M., 1979. Population biology of infectious diseases: part I. Nature 280, 361–367.

Anderson, R.M., May, R.M., 1991. Infectious Diseases of Humans. Dynamics and Control. Oxford University Press, Oxford.

Bartlett, M.S., 1960. The critical community size for measles in the United States. J. R. Stat. Soc. A 123, 37–44.

Bartlett, M.S., 1961. Monte Carlo studies in ecology and epidemiology. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, CA, Vol. 4, pp. 39–55.

Bjørnstad, O.N., Finkenstädt, B.F., Grenfell, B.T., 2002. Dynamics of measles epidemics. I. Estimating scaling of transmission rates using a time series SIR model. Ecol. Monogr. 72, 169–184.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J.R. Stat. Soc. Ser. B 26, 211–252.

Earn, D.J.D., Rohani, P., Bolker, B.M., Grenfell, B.T., 2000. A simple model for complex dynamical transitions in epidemics. Science 287, 667–670.

Fine, P.E.M., Clarkson, J.A., 1982. Measles in England and Wales. I. An analysis of factors underlying seasonal patterns. Int. J. Epidemiol. 11, 5–14.

Finkenstädt, B.F., Grenfell, B.T., 2000. Time series modelling of childhood diseases: a dynamical systems approach. Appl. Stat. 49 (2), 187–205.

Godfray, H.C.J., Grenfell, B.T., 1993. The continuing quest for chaos. Trends Ecol. Evol. 8 (2), 43–44.

Grenfell, B., Harwood, J., 1997. (Meta)population dynamics of infectious diseases. Trends Ecol. Evol. 12, 395–399.

Grenfell, B.T., Bjørnstad, O.N., Kappey, J., 2001. Travelling waves and spatial hierarchies in measles epidemics. Nature 414, 716–723.

Grenfell, B.T., Bjørnstad, O.N., Finkenstädt, B.F., 2002. Dynamics of measles epidemics. II. Scaling noise, determinism and predictability with the time series SIR model. Ecol. Monogr. 72, 185–202.

Keeling, M.J., Grenfell, B.T., 1997. Disease extinction and community size: modelling the persistence of measles. Science 275, 65–67.

Keeling, M.J., Rohani, P., Grenfell, B.T., 2001. Seasonally forced disease dynamics explored as switching between attractors. Physica D 148 (3–4), 317–335.

Mollison, D., Din, S.U., 1993. Deterministic and stochastic models for the seasonal variability of measles transmission. Math. Biosci. 117, 155–177.

Mollison, D., Isham, V., Grenfell, B., 1994. Epidemics: models and data. J. R. Stat. Soc. A 157, 115–149.

Murray, J.D., 1989. Mathematical Biology. Springer, Berlin.

Nicholson, A.J., Bailey, V.A., 1935. The balance of animal populations—Part 1. Proc. Zool. Soc. London 1, 551–598.

Schenzle, D., 1984. An age-structured model of pre- and post-vaccination measles transmission. IMA J. Math. Appl. Med. Biol. 1, 169–191.