

## Effects of the infectious period distribution on predicted transitions in childhood disease dynamics

Olga Krylova and David J. D. Earn

*J. R. Soc. Interface* 2013 **10**, 20130098, published 15 May 2013

---

### Supplementary data

["Data Supplement"](#)

<http://rsif.royalsocietypublishing.org/content/suppl/2013/05/09/rsif.2013.0098.DC1.html>

### References

[This article cites 51 articles, 20 of which can be accessed free](#)

<http://rsif.royalsocietypublishing.org/content/10/84/20130098.full.html#ref-list-1>

### Subject collections

Articles on similar topics can be found in the following collections

[biomathematics](#) (221 articles)

### Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)



## Research

**Cite this article:** Krylova O, Earn DJD. 2013 Effects of the infectious period distribution on predicted transitions in childhood disease dynamics. *J R Soc Interface* 10: 20130098. <http://dx.doi.org/10.1098/rsif.2013.0098>

Received: 31 January 2013

Accepted: 18 April 2013

### Subject Areas:

biomathematics

### Keywords:

SIR epidemic model, seasonal forcing, waiting time distribution, bifurcation theory, generation time, measles in New York City

### Author for correspondence:

David J. D. Earn

e-mail: [earn@math.mcmaster.ca](mailto:earn@math.mcmaster.ca)

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2013.0098> or via <http://rsif.royalsocietypublishing.org>.

# Effects of the infectious period distribution on predicted transitions in childhood disease dynamics

Olga Krylova and David J. D. Earn

Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada L8S 4K1

The population dynamics of infectious diseases occasionally undergo rapid qualitative changes, such as transitions from annual to biennial cycles or to irregular dynamics. Previous work, based on the standard seasonally forced ‘susceptible–exposed–infectious–removed’ (SEIR) model has found that transitions in the dynamics of many childhood diseases result from bifurcations induced by slow changes in birth and vaccination rates. However, the standard SEIR formulation assumes that the stage durations (latent and infectious periods) are exponentially distributed, whereas real distributions are narrower and centred around the mean. Much recent work has indicated that realistically distributed stage durations strongly affect the dynamical structure of seasonally forced epidemic models. We investigate whether inferences drawn from previous analyses of transitions in patterns of measles dynamics are robust to the shapes of the stage duration distributions. As an illustrative example, we analyse measles dynamics in New York City from 1928 to 1972. We find that with a fixed mean infectious period in the susceptible–infectious–removed (SIR) model, the dynamical structure and predicted transitions vary substantially as a function of the shape of the infectious period distribution. By contrast, with fixed mean latent and infectious periods in the SEIR model, the shapes of the stage duration distributions have a less dramatic effect on model dynamical structure and predicted transitions. All these results can be understood more easily by considering the distribution of the disease generation time as opposed to the distributions of individual disease stages. Numerical bifurcation analysis reveals that for a given mean generation time the dynamics of the SIR and SEIR models for measles are nearly equivalent and are insensitive to the shapes of the disease stage distributions.

## 1 Introduction

Mathematical modelling has proven to be an extremely powerful tool for understanding epidemiological patterns and predicting how demographic changes and control measures influence infectious disease dynamics [1–3]. The most commonly used framework for modelling transmission dynamics involves dividing the population into compartments based on disease status and using ordinary differential equations (ODEs) to specify flows between the compartments. For diseases that confer permanent immunity, the simplest case is the SIR model [1,4], in which the compartments represent susceptible, infectious and removed individuals, while the SEIR model also includes an exposed compartment, containing individuals who are in a latent stage (infected but not yet infectious). These simple models implicitly assume that the time an individual spends in each disease stage (e.g. latent or infectious) is drawn from exponential distributions [2,5], which are unlike real distributions of disease stage durations.

The dynamical effects of exponential versus more realistic distributions of stage durations have been explored extensively in the literature [6–12], which has revealed that changing the shapes of these distributions while keeping their means fixed can have a large impact on predicted dynamics. Consequently, it is important to re-evaluate any inferences drawn about real data

from models that assume exponentially distributed stage durations. In this paper, we study the dynamics of a family of SIR and SEIR models with stage duration distributions that range from exponential, to realistically bell-shaped, to fixed. We investigate how the shapes of latent and infectious period distributions affect our predictions concerning epidemiological transitions (e.g. from annual to biennial epidemic cycles) and compare our results with conclusions previously made based on exponentially distributed models [13–15]. As an illustrative example, we apply our analysis to measles epidemics in New York City from 1928 to 1972.

### 1.1. The shapes of real distributions of disease stage durations

Many authors have estimated infectious period distributions by fitting standard probability distributions (e.g. normal [16–18], log-normal [19,20], gamma [9,21] or fixed [16,17]) to empirical data. For transmission modelling, a gamma distribution with an integer shape parameter—also known as an *Erlang distribution*—is strongly preferred on theoretical grounds: the Erlang distribution is equivalent to a sequence of independent and identically distributed exponential distributions [6,22–24], so compartmental transmission models with Erlang-distributed stage durations can be expressed as ODEs (as opposed to the integro-differential equations required to express compartmental models with arbitrarily distributed stage durations).

The Erlang distribution with shape parameter  $n$  and scale parameter  $n\gamma$ , Erlang( $n, n\gamma$ ), has probability density

$$f(x; n, n\gamma) = \frac{(n\gamma)^n}{(n-1)!} x^{(n-1)} e^{-n\gamma x}, \quad x > 0, n \in \mathbb{N}. \quad (1.1)$$

The mean is  $1/\gamma$  and the variance is  $1/n\gamma^2$ .

The Erlang distribution is more restricted in shape than the general gamma distribution, but it is sufficiently flexible to provide a good approximation of realistic stage duration distributions. Figure 1 shows the probability density function of the Erlang distribution with mean  $1/\gamma = 13$  days (vertical line) and various shape parameters ( $n = 1, 2, 3, 5, 8, 20, 100$ ).

We write  $SI^mR$  and  $SE^nI^mR$  to refer to the Erlang-distributed SIR and SEIR models, where  $m$  and  $n$  refer to the shape parameters of the latent and infectious period distributions, respectively. Thus,  $SI^1R$  ( $n = 1$ ) and  $SE^1I^1R$  ( $m = 1, n = 1$ ) denote the standard SIR and SEIR models with exponentially distributed latent and infectious periods. Estimated values of  $n$  and  $m$  can be inferred from appropriate clinical data and vary widely for different infectious diseases, for example,  $m = 2, n = 3$  for SARS and  $m = 20, n = 20$  for measles [9].

### 1.2. The Erlang-distributed epidemic models

In our analysis, we use standard Erlang-distributed SIR (equation (1.2)) and SEIR models [6,8–10,23,24].

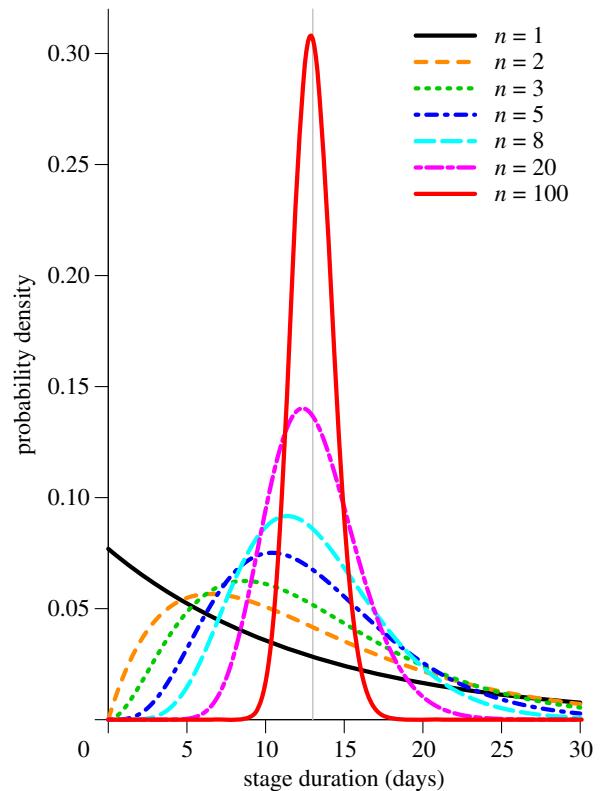
$$\frac{dS}{dt} = \nu N_0 - \beta SI - \mu S, \quad (1.2a)$$

$$\frac{dI_1}{dt} = \beta SI - (n\gamma + \mu)I_1, \quad (1.2b)$$

$$\frac{dI_2}{dt} = n\gamma I_1 - (n\gamma + \mu)I_2 \quad (1.2c)$$

⋮

and 
$$\frac{dI_n}{dt} = n\gamma I_{n-1} - (n\gamma + \mu)I_n. \quad (1.2d)$$



**Figure 1.** Probability density functions for several Erlang distributions with the same mean (13 days, marked with a vertical grey line) but different shape parameter  $n$  (see equation (1.1)). The most extreme cases are the exponential distribution ( $n = 1$ ) and the Dirac delta distribution ( $n \rightarrow \infty$ ). (Online version in colour.)

Here,  $S$ ,  $I$  and  $R$  are the numbers of susceptible, infectious and recovered (immune) individuals in the population.  $\mu$ ,  $\beta$  and  $\gamma$  are the rates of *per capita* death, transmission and recovery, respectively.  $\mu$  quantifies death from ‘natural causes’ (disease-induced mortality is assumed to be negligible).  $\beta$  is the rate at which contacts between susceptible and infectious individuals cause new infections (per susceptible per infected). The term  $\nu N_0$  denotes the number of births per unit time, where  $N_0$  is the population size at a particular ‘anchor time’  $t_0$  and  $\nu$  represents births *per capita* at time  $t_0$ , but not at other times (see also §§2.1 and 2.2 and electronic supplementary material, section ‘Models’). This term is particularly important, because secular changes in this birth rate can induce dynamical transitions [3,13–15]. In our formulation, the birth term ( $\nu N_0$ ) is different from the birth term in typical SIR-based model formulation, which assumes that births balance deaths with birth rate being  $\mu N$ . We estimate  $\nu N_0$  based on demographic data and do not assume that it scales with population size (e.g. we do *not* assume that the birth rate is  $\mu N$ ).

In equation (1.2), the infectious stage is broken up into a sequence of  $n$  substages, each exponentially distributed with mean  $1/(n\gamma)$ . The full infectious period distribution is the Erlang distribution with shape parameter  $n$  and scale parameter  $n\gamma$ , Erlang( $n, n\gamma$ ).

Transmission of childhood diseases such as measles is strongly influenced by seasonal changes in contact rates among children [13,25]. We assume that the transmission rate varies sinusoidally over the course of a year,

$$\beta(t) = \langle \beta \rangle (1 + \alpha \cos(2\pi t)), \quad (1.3)$$

where  $\langle\beta\rangle$  is the mean transmission rate and  $\alpha$  is the amplitude of seasonal forcing. (See the electronic supplementary material, section ‘Models’.)

A fundamental characteristic of an infectious disease is its *basic reproduction number*,  $\mathcal{R}_0$ , which is the mean number of susceptible individuals infected by one infectious individual in a completely susceptible population [1]. Formally defining and interpreting  $\mathcal{R}_0$  in the presence of periodic forcing of parameters requires considerable mathematical care [26,27]; however, what is important for our purposes here is that the threshold for disease spread is determined by the more easily defined basic reproduction number for the time-averaged system [28]—i.e. the autonomous system in which  $\beta(t)$  is replaced by  $\langle\beta\rangle$ —and this is what we shall always mean when referring to ‘ $\mathcal{R}_0$ ’. Thus,  $\mathcal{R}_0$  is the product of the mean transmission rate  $\langle\beta\rangle$  (cf. equation (1.3)) and the mean duration of infectiousness  $T_{\text{inf}}$  and an epidemic can occur only if  $\mathcal{R}_0 > 1$ . The exact expression for  $T_{\text{inf}}$  for Erlang-distributed models is cumbersome (see the electronic supplementary material, section ‘Models’) but for typical respiratory infections—for which the duration of infection is much shorter than the average host lifetime  $1/\mu$ —it is always true that  $T_{\text{inf}} \approx 1/\gamma$  and hence

$$\mathcal{R}_0 \approx \frac{\nu N_0 \langle\beta\rangle}{\mu \gamma}. \quad (1.4)$$

The first factor here ( $\nu N_0/\mu$ ) does not normally appear in formulae for  $\mathcal{R}_0$  because it is typically assumed that births balance deaths, and the population size is often absorbed into the transmission rate  $\beta$  (see the ‘Models’ section of the electronic supplementary material for a more formal discussion of this point). We assume that  $\nu$  changes slowly enough that it can be regarded as constant for the purposes of defining  $\mathcal{R}_0$  at a given time.

Detailed descriptions of the  $S^mI^mR$  and  $SE^mI^mR$  models can be found in the electronic supplementary material, section ‘Models’.

### 1.3. Dynamics of epidemic models with Erlang-distributed stage durations

In the past 20 years, the  $S^mI^mR$  and  $SE^mI^mR$  models—and other more general models—have received a great deal of attention. Equilibrium stability analyses have been conducted on ‘unforced’ models that assume constant contact rates [6,7,29–32], and bifurcation analyses have been conducted on ‘forced’ models in which contact rates vary seasonally [6–12,33]. Lloyd [7] found that the biennial pattern observed in the  $S^1I^1R$  model is reproduced by the  $S^mI^mR$  model but with much weaker seasonality. Nguyen & Rohani [10] found that complex dynamics of whooping cough could be understood based on the multiple coexisting attractors of an  $SE^1I^1R$  model, whereas the simple  $SE^1I^1R$  model with the same mean latent and infectious periods always predicts an asymptotically annual cycle. Wearing *et al.* [9] argued that the traditional assumptions of exponentially distributed latent and infectious periods may lead to underestimation of the basic reproduction number,  $\mathcal{R}_0$ , and hence to underestimation of the levels of control required to curtail an epidemic.

The primary theme of recent work on  $S^mI^mR$  and  $SE^mI^mR$  models has been that the shapes of stage duration distributions can significantly affect the qualitative dynamics of infectious diseases. Given this, it is important to re-examine previous work that has attempted to explain observed disease

dynamics based on  $S^1I^1R$  or  $SE^1I^1R$  models, and determine whether the conclusions of these previous studies remain valid when the analyses are repeated using models with more realistically distributed stage durations. Our particular focus in this paper is on epidemiological transition analysis, by which we mean predicting qualitative changes in epidemic dynamics induced by demographic and behavioural changes in the host population [3,13,15]. As an illustrative example, we analyse measles incidence in New York City for the period 1928–1972, which was first investigated by London & Yorke [25,34] and has been the subject of numerous studies over the past 40 years [13,15,35,36]. We also investigate whether the dynamics of a given  $SE^mI^mR$  model can be approximated with an  $S^mI^mR$  model.

We begin by describing the method of transition analysis in §2. In §3, we apply transition analysis, based on  $S^mI^mR$  and  $SE^mI^mR$  models, to measles dynamics in New York City from 1928 to 1972. We consider the role of the distribution of the disease generation time (as opposed to the latent and infectious periods) in §4 and summarize our results in §5.

## 2. Predicting epidemiological transitions

Many infectious disease time series display occasional, rapid changes in qualitative dynamics, such as transitions from annual to biennial cycles or to irregular dynamics [1,35]. Previous work has shown that these transitions appear to be driven by demographic and behavioural changes that induce bifurcations in the  $SE^1I^1R$  model [3,13,15]. We would like to know whether the qualitative inferences made previously based on the  $SE^1I^1R$  model remain valid when the analysis is repeated with more realistic  $SE^mI^mR$  models.

Earn *et al.* [13] used the  $SE^1I^1R$  model to show that knowing the changes in birth and vaccination rates—or, more generally, changes in the rate at which susceptible individuals are recruited into the population—it is possible to predict the occurrence of bifurcations that change the period of epidemic cycles. We briefly revisit that argument here in the more general context of the  $S^mI^mR$  model.

### 2.1. Theoretical motivation for transition analysis

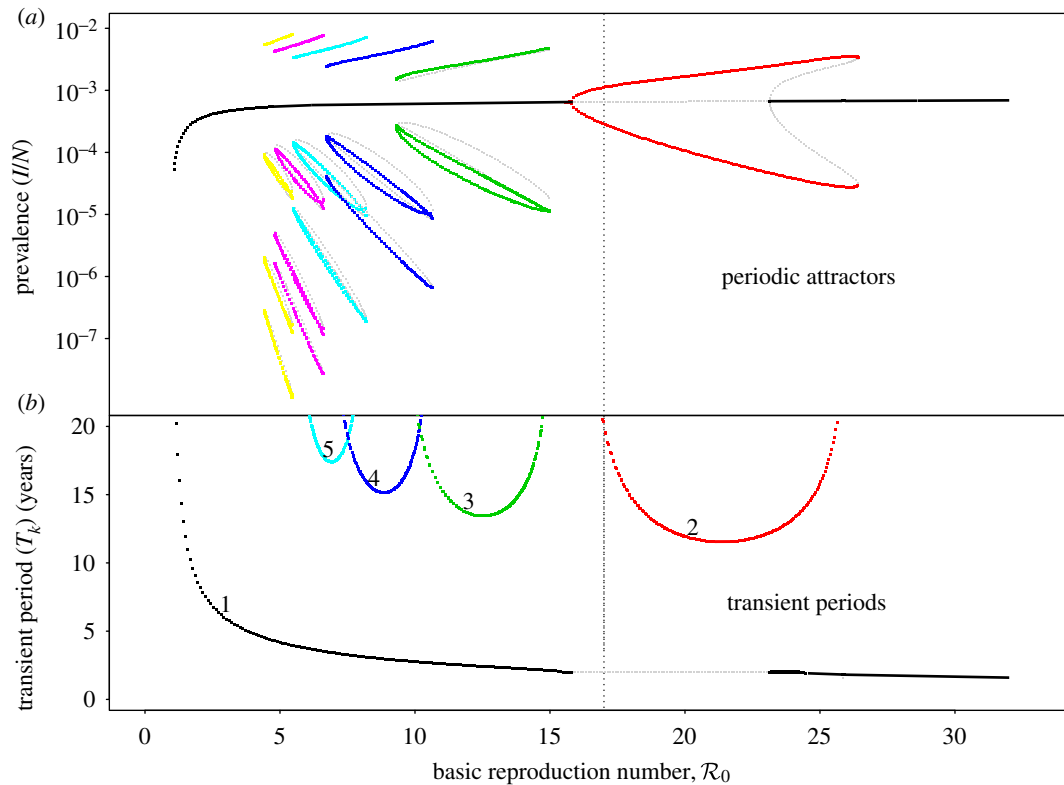
In equation (1.2a), the factor  $\nu$  was formulated as the birth rate but can be thought of more generally as the susceptible recruitment rate. Suppose that this rate changes to  $\nu'$ , which might occur because the birth rate has changed or because we have begun to vaccinate a proportion  $p$  of the population (in which case  $\nu' = \nu(1 - p)$ ). To understand the dynamical effect of this change from  $\nu$  to  $\nu'$ , consider the following simple change of variables:

$$S' = \frac{\nu}{\nu'} S, \quad I'_k = \frac{\nu}{\nu'} I_k, \quad \text{for } 1 \leq k \leq n. \quad (2.1)$$

If we insert these expressions in equation (1.2) and solve the equations for the primed variables we obtain, for example,

$$\frac{dS'}{dt} = \nu N_0 - \beta \frac{\nu'}{\nu} S' I' - \mu S'. \quad (2.2)$$

That is, the equations for the primed variables are identical to the original equations (with the original susceptible recruitment term  $\nu$ ), but with the transmission rate changed from  $\beta$  to  $\beta\nu'/\nu$ . Thus, the dynamical effect of a change in susceptible recruitment by a given factor is identical to the



**Figure 2.** Asymptotic and perturbation analysis of the sinusoidally forced  $SI^1R$  model (equations (1.2),  $n = 1$ ) parameterized for measles ( $\gamma^{-1} = 13$  days,  $\nu = 0.02 \text{ yr}^{-1}$ ,  $\alpha = 0.08$ ). (a) Asymptotic analysis: the bifurcation diagram for the model with control parameter  $\mathcal{R}_0$ . The ordinate shows the proportional prevalence of infection at the start of each year, so annual cycles are indicated by a single point at each  $\mathcal{R}_0$ , biennial cycles by two points, triennial cycles by three, and so on. Heavy curves correspond to stable cycles while light curves indicate unstable cycles. A dotted vertical line is drawn at  $\mathcal{R}_0 = 17$ , indicating the estimate of the basic reproduction number at the ‘anchor time’  $t_0$ . Two types of bifurcations occur in this diagram: period doublings (also called pitchforks or flips) and tangent bifurcations (also called folds or saddle–node bifurcations). (b) Perturbation analysis: the natural period of damped oscillations (the transient period) onto each attractor, as described in step 2 of §2.2. The transient period curves are labelled according to the corresponding attractor in the (a): transient period of the annual attractor (1), biennial attractor (2), triennial (3) and so on. The light line indicates a region where the annual cycle is unstable and the period of repelled transients is phase-locked at exactly 2 years [37]. (Online version in colour.)

dynamical effect of changing the transmission rate by exactly that factor,

$$\nu \rightarrow \nu' \Rightarrow \beta \rightarrow \beta \frac{\nu'}{\nu} \quad \text{and} \quad \mathcal{R}_0 \rightarrow \mathcal{R}_0 \frac{\nu'}{\nu}. \quad (2.3)$$

Consequently, we can use a bifurcation diagram with the transmission rate  $\beta$ , or equivalently the basic reproduction number  $\mathcal{R}_0$  (because  $\mathcal{R}_0$  is proportional to  $\beta$ ), as the control parameter to predict transitions in dynamical behaviour induced by changes in susceptible recruitment rate. Figure 2 shows such a bifurcation diagram based on the sinusoidally forced  $SI^1R$  model (equation (1.2),  $n = 1$ ) with parameters chosen to correspond to measles (and with an estimated value of  $\mathcal{R}_0 = 17$  at some given time, say  $t_0$ , marked with a dotted vertical line). If the susceptible recruitment rate was  $\nu_0$  at time  $t_0$  and  $\nu_1$  at time  $t_1$ , then we would predict that at time  $t_1$  the system would behave as if the basic reproduction number had changed by the factor  $\nu_1/\nu_0$ , i.e. the *effective* reproduction number at time  $t$ , is

$$\mathcal{R}_{0,\text{eff}} = \mathcal{R}_0 \frac{\nu_1}{\nu_0}. \quad (2.4)$$

There is an important subtlety upon which our ability to predict transitions depends critically. In the equation for  $dS/dt$  (equations (1.2a)), the susceptible recruitment rate appears as a constant ( $\nu$  does not depend explicitly on time  $t$  or population size  $N$ ), and we use mass-action incidence

( $\beta SI$ ) rather than standard incidence ( $\beta SI/N$ ). If the susceptible recruitment term were taken to be  $\nu N$  rather than  $\nu N_0$ , and we were to use standard incidence then the variable change in equation (2.1) would have no effect (the differential equations are invariant to the scaling transformation given by equation (2.1)) and we would never predict dynamical transitions resulting from changes in the susceptible recruitment rate. One can debate on theoretical grounds whether one model formulation or another is most plausible biologically [38]; we favour our formulation because it leads to correct predictions concerning dynamical transitions [13,15]. We are interested in the effects of changes in  $\nu$  over time, but the changes of interest occur slowly compared with the epidemic timescale, which is why we can treat  $\nu$  as constant in the  $dS/dt$  equation.

## 2.2. The method of transition analysis

Given a time series of reported disease incidence or mortality (for a disease for which we have estimates of the mean latent and infectious periods), a full *transition analysis* proceeds as follows [13,15]. First, in order to clarify what needs to be explained, plot the disease time series together with its estimated frequency structure at each time point (e.g. Fourier power spectra for subsets of the full time series or, preferably, a wavelet spectrum for the full time series [39,40]). Second, for some ‘anchor time’  $t_0$  in the time series, obtain an estimate

of the basic reproduction number  $\mathcal{R}_0$ , preferably using data other than the focal time series (e.g. annual age-specific data [1]). Third, estimate the susceptible recruitment rate  $\nu$  at each point of the disease time series and infer the effective reproductive number  $\mathcal{R}_{0,\text{eff}}$  at all times by inserting the estimated  $\nu$  values into equation (2.4) (where  $\nu_0 = \nu(t_0)$  and  $\nu_1 = \nu(t)$  for an arbitrary time  $t$ ). Fourth, identify time intervals during which  $\nu$  is roughly constant (hence during which the dynamical features of the disease time series can be expected to be approximately stationary). Finally, based on the estimated value of  $\mathcal{R}_{0,\text{eff}}$  in each of the ‘dynamically stationary time intervals’, predict transitions in qualitative dynamical behaviour (e.g. changes in the structure of the wavelet spectrum, especially the positions of peaks), as follows.

- (1) *Asymptotic analysis* (to identify the periods of *attractors* of the model, which are reached asymptotically) [10–15]: construct a bifurcation diagram with  $\mathcal{R}_0$  as the control parameter, over a range of  $\mathcal{R}_0$  that includes the value estimated for time  $t_0$  and the full range of  $\mathcal{R}_{0,\text{eff}}$  determined via equation (2.4) (figure 2a). From this diagram, we can easily infer the periods of cyclical attractors of the system. We call these *resonant periods* because they are exact subharmonics (i.e. integer multiples) of the period of seasonal forcing (1 year). (See the electronic supplementary material ‘Bifurcation analysis of the seasonally forced SIR model using XPPAUT’ for a step-by-step guide to creating diagrams such as figure 2 using XPPAUT [41].)
- (2) *Perturbation analysis* (to estimate the periods of the *transients* associated with each attractor): over the same range of  $\mathcal{R}_0$  as in the asymptotic analysis, plot the periods of the transients associated with—i.e. the periods of damped oscillations onto—each cyclical attractor (figure 2b). We call these *non-resonant periods* because they can take any real value and are not entrained by seasonal forcing. Non-resonant periods may be detected in observed epidemic time series, because transients can be sustained by demographic stochasticity [15,42]. Non-resonant periods can be calculated by linearizing about the fixed points and cycles of the model’s 1-year-stroboscopic map [14,15]. If the period of a given attractor is  $k$  and the dominant eigenvalue of the associated  $k$ -cycle of the stroboscopic map is  $\lambda k$  (which is complex for typical disease parameters), then the associated transient period is

$$T_k = \frac{2\pi k}{|\text{Arg}(\lambda_k)|}. \quad (2.5)$$

- (3) *Stochastic analysis* (to estimate the relative importance of transient versus asymptotic dynamics): the wavelet spectrum has peaks at the most important periods in the time series (which we attempt to predict with steps 1 and 2) but also shows the magnitude of the peaks, which cannot be estimated by asymptotic and perturbation analysis of a deterministic model. The relative magnitudes of spectral peaks of observed time series can be estimated from spectra of simulations of stochastic realizations of the model, with the expectation that smaller population sizes (which are subject to greater demographic stochasticity) will stimulate more transient dynamics, leading to larger spectral peaks at non-resonant periods [3,15]. Because the stochastic analysis addresses the details rather than the main features of dynamical transitions, we do not conduct it in this

paper (though we make occasional reference to stochastic effects). We note, however, that understanding these details is an area of very active research, and powerful analytical approaches for estimating power spectra for recurrent epidemic processes have been developed recently [11,43–45]. Ultimately, a complete transition theory would need to account for all the dynamical characteristics of stochastic epidemic models, which include alternation between asymptotic and transient behaviour [15], switching between different attractors [13,46], phase-locked cycles at one fixed period [37] and interactions with repellers [47].

In §3, we use the  $\text{SI}^n\text{R}$  and  $\text{SE}^m\text{I}^n\text{R}$  models to conduct transition analysis of the well-known New York City measles time series [34]. Our main question is: do we predict different transitions if we base our theoretical analysis on the  $\text{SI}^n\text{R}$  rather than on the  $\text{SI}^1\text{R}$  model, or the  $\text{SE}^m\text{I}^n\text{R}$  rather than  $\text{SE}^1\text{I}^1\text{R}$  model?

Another question that we will address is: can we approximate the dynamics of the  $\text{SE}^m\text{I}^n\text{R}$  model using the  $\text{SI}^n\text{R}$  model? This question is motivated by the fact that the dynamics of the  $\text{SE}^1\text{I}^1\text{R}$  model can be approximated using the  $\text{SI}^1\text{R}$  model. It is well-known that the equilibrium and stability properties (e.g. the period of damped oscillations onto the equilibrium) of the unforced  $\text{SI}^1\text{R}$  and  $\text{SE}^1\text{I}^1\text{R}$  models correspond if the mean infectious period in the  $\text{SI}^1\text{R}$  model is associated with the sum of the mean latent and mean infectious periods in the  $\text{SE}^1\text{I}^1\text{R}$  model [1, p. 668]. The measles bifurcation diagram shown in figure 2 for the sinusoidally forced  $\text{SI}^1\text{R}$  model is virtually identical to the term-time forced  $\text{SE}^1\text{I}^1\text{R}$  measles bifurcation diagram produced previously by Earn *et al.* [13]. Therefore, we analyse the  $\text{SI}^n\text{R}$  model with mean infectious period  $1/\gamma = 13$  days (the sum of the real mean latent period of 8 days and the real mean infectious period of 5 days for measles).

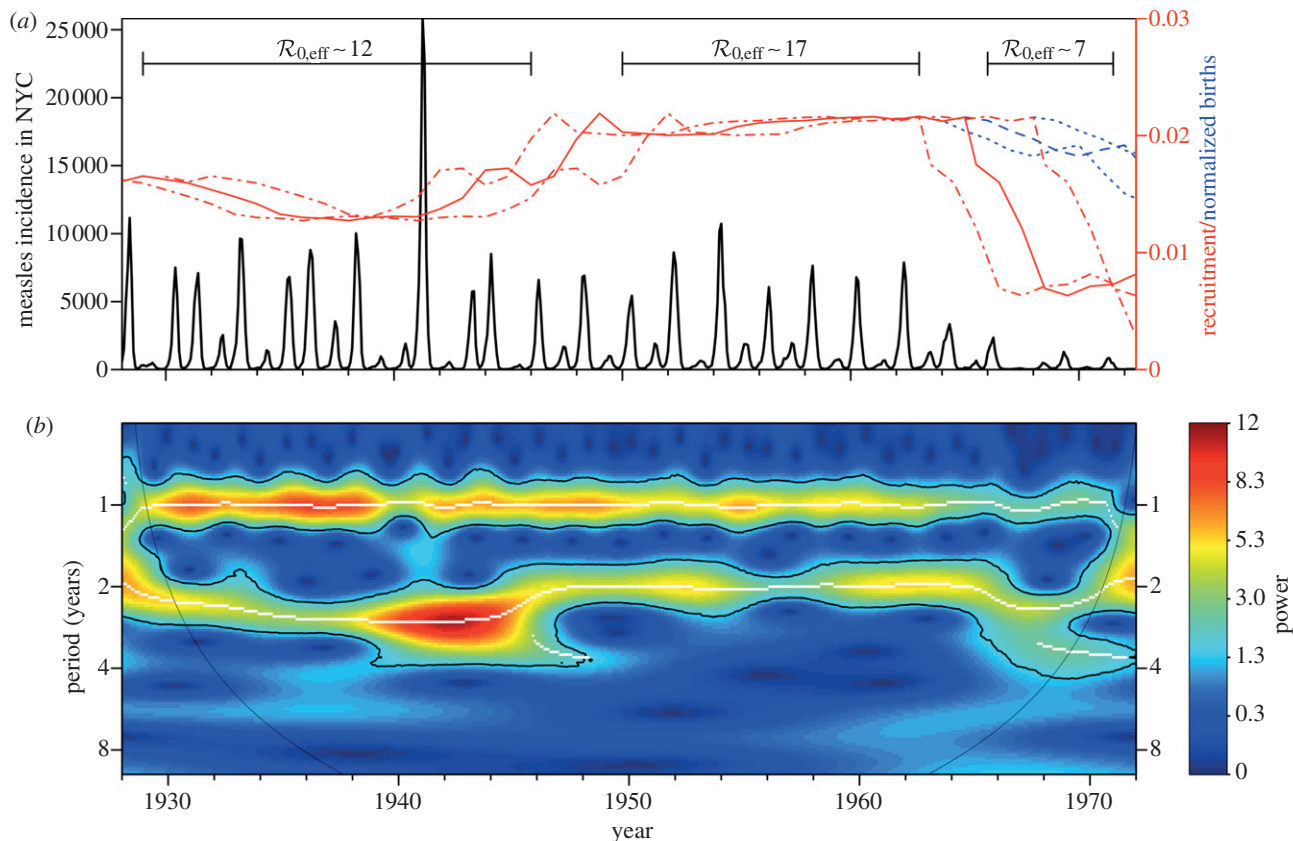
### 3. Transition analysis using $\text{SI}^n\text{R}$ and $\text{SE}^m\text{I}^n\text{R}$ models

In this section, we use the well-known measles incidence time series for New York City (1928–1972) as an illustrative example with which to compare the results of transition analysis using  $\text{SI}^n\text{R}$  and  $\text{SE}^m\text{I}^n\text{R}$  models with stage duration distributions varying from exponential to fixed. The New York City measles data were originally digitized and studied by London & Yorke [25,34]. Previous transition analysis of these data [13,15] has been restricted to the pre-vaccine period (up to 1963). Here, we are able to extend our analysis to 1972 using vaccination data for 1963–1972 (see the electronic supplementary material, section ‘Vaccination level calculations’).

#### 3.1. Description of the data

##### 3.1.1. Reported incidence and inferred frequency structure

Figure 3a shows monthly reported cases of measles in New York City (together with estimated susceptible recruitment rate) and figure 3b shows the frequency structure of the data over time as a wavelet spectrum. Two spectral peaks are evident for the full duration of the time series, one at a period of 1 year and a second at a period that



**Figure 3.** Measles in New York City, 1928–1972. (a) Monthly reported measles cases (heavy solid oscillating curve) and annual susceptible recruitment relative to the population size in 1960 (upper solid curve). Before the introduction of vaccination in 1963, annual susceptible recruitment coincided with annual births (dashed). The susceptible recruitment curve is shifted forward by 2 years to account for the delay between birth and entering the well-mixed population. Dotted-dashed curves show the susceptible recruitment rate without delay (2 years earlier than the solid curve) and with a delay of 5 years (3 years later than the solid curve). Similarly, the dotted curves after 1963 show the birth rate without delay and with a 5-year delay. The line segments at the top of panel (a) highlight time intervals with distinct effective  $\mathcal{R}_0$ , estimated with equation (2.4). (b) The wavelet power spectrum of the measles incidence time series (log-transformed and normalized to unit variance). The white curves show the local maxima of wavelet power (squared modulus of wavelet coefficients [48, p. 291]) at each time. The black curves indicate 95% confidence regions, estimated from 1000 bootstrapped time series [48, pp. 292–293]. Below the ‘cone of influence’ [48,49], the calculation of wavelet power is less accurate because it includes edges of the time series that have been zero-padded to make the length of the series a power of 2. The wavelet spectrum was computed using MATLAB code kindly provided by Bernard Cazelles [48–50]. (Online version in colour.)

changes over time (2–3 years from 1928 to about 1946, exactly 2 years from about 1946 to 1965, and 2–4 years from about 1965 to the end of 1972).

### 3.1.2. Estimated susceptible recruitment

Based on age-incidence and age-seroprevalence data for England and Wales (1950–1968), the basic reproduction number for measles has been estimated to be  $\mathcal{R}_0 \simeq 17$  in the pre-vaccination era [1, fig. 3.9 and 3.10, and table 4.1, p. 70]. Because, in New York City, the birth rate was approximately the same as in England and Wales (in the pre-vaccination era), we use this value as an estimate for  $\mathcal{R}_0$  in New York City in 1960, which we take to be our ‘anchor time’  $t_0$ .

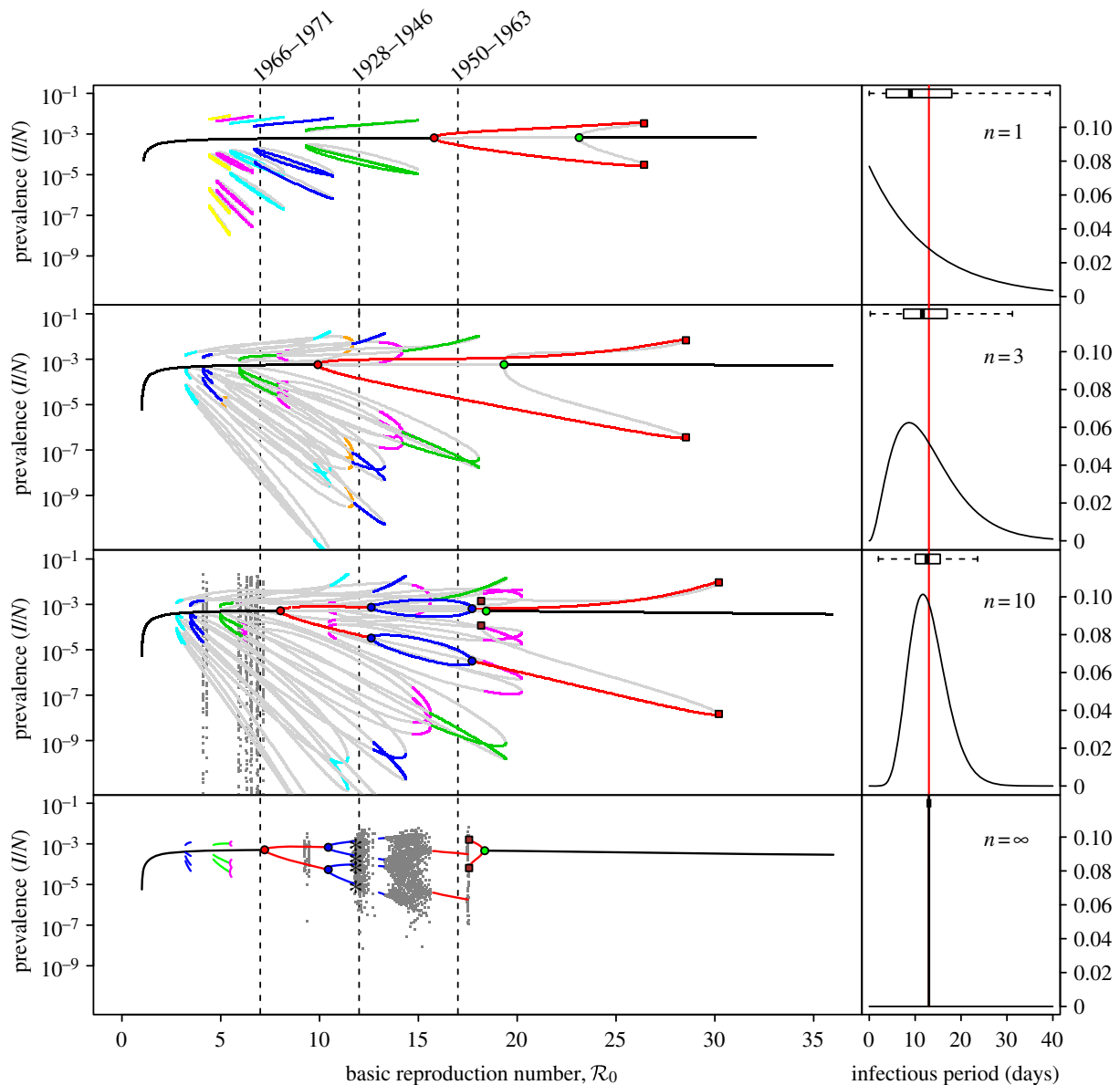
Measles vaccine was introduced in the United States in 1963 [51], so susceptible recruitment until 1963 can be taken to be associated entirely with births. However, newborns do not enter the well-mixed susceptible pool immediately, for two reasons: (i) maternally acquired immunity can take up to a year to wane [1, p. 50], (ii) before entering pre-school, children typically have much lower contact rates with other susceptibles. Hence, the impact of changes in birth rate on transmission dynamics is delayed, approximately by the time between birth and entering the well-mixed susceptible pool. We took this delay,  $\tau_S$ , to be 2

years, but our conclusions are not sensitive to this parameter (e.g. taking it to be 0 or 5 years makes little difference (dotted-dashed and dotted curves in figure 3)). Note that  $\tau_S$  should be less than 5 because the mean age at infection was about 5 years [1, fig. 8.1, p. 156]. Thus, we take the susceptible recruitment rate in 1960 to be the ratio of the number of births in 1958 ( $B(t_0 - \tau_S) = 167\,660$ ) to the estimated population of New York City in 1960 ( $N_0 = 7\,781\,984$ ), i.e.  $\nu(t_0) \simeq 0.02$  [52]. At other times  $t$ ,

$$\nu(t) = \frac{B(t - \tau_S)}{N_0} (1 - p(t - \tau_S)), \quad (3.1)$$

where  $p(t)$  is the proportion of new recruits at time  $t$  who were vaccinated before entering the well-mixed susceptible pool. Note in equation (3.1) we use  $N_0$ , not  $N(t)$ : recruitment is normalized relative to the population size at the ‘anchor time’  $t_0$  [13]. After 1963, the susceptible recruitment rate is substantially reduced by the introduction of vaccination (figure 3).

The birth and measles vaccination data that we insert in equation (3.1) are discussed in the electronic supplementary material, section ‘Vaccination level calculations’. The resulting annual susceptible recruitment rate is shown in figure 3a. There are three distinct periods during which the recruitment rate was roughly constant: 1929–1946 with  $\nu \approx 0.015$ , 1950–1963 with  $\nu \approx 0.02$  and 1966–1971 with  $\nu \approx 0.008$ .



**Figure 4.**  $SI^nR$  measles bifurcation diagrams as a function of  $\mathcal{R}_0$  for several values of the shape parameter of the infectious period distribution ( $n = 1, 3, 10, \infty$ ), with other parameters fixed (mean infectious period  $1/\gamma = 13$  days, birth rate  $\nu = 0.02$  per year, seasonal forcing amplitude  $\alpha = 0.08$ ). Heavy curves show attractors. In the online version, attractors of different periods are drawn in different colours. Light grey curves show unstable branches. Circles represent period-doubling (flip) bifurcations while squares denote tangent (saddle–node) bifurcations of the main branch. Dashed vertical lines highlight  $\mathcal{R}_{0,\text{eff}} = 7, 12$  and  $17$ , which correspond to the estimated effective reproduction number for measles in New York City for the year ranges indicated, as in figure 3. Each right panel shows the corresponding probability distribution of the infectious period and a box plot showing the 5%, 25%, 50%, 75% and 95% quantiles of the distribution; a vertical line shows the mean infectious period (13 days). For finite  $n$ , the bifurcation diagrams were computed using standard continuation software (XPPAUT [41]), whereas the fixed-delay limit ( $n = \infty$ ) was computed by ‘brute force’, i.e. by numerical integration of the delay differential equation (see the electronic supplementary material, equation S12) until convergence onto an attractor; hence, unstable branches are not shown in the limit  $n = \infty$ . Brute force bifurcation diagrams were also computed for finite  $n$  to reveal regions of chaotic behaviour which are shown in grey. (Online version in colour.)

Therefore, from equation (2.4), we estimate the effective reproduction number to be  $\mathcal{R}_{0,\text{eff}} \approx 12$  for 1928–1946,  $\mathcal{R}_{0,\text{eff}} \approx 17$  for 1950–1963 and  $\mathcal{R}_{0,\text{eff}} \approx 7$  for 1966–1971.

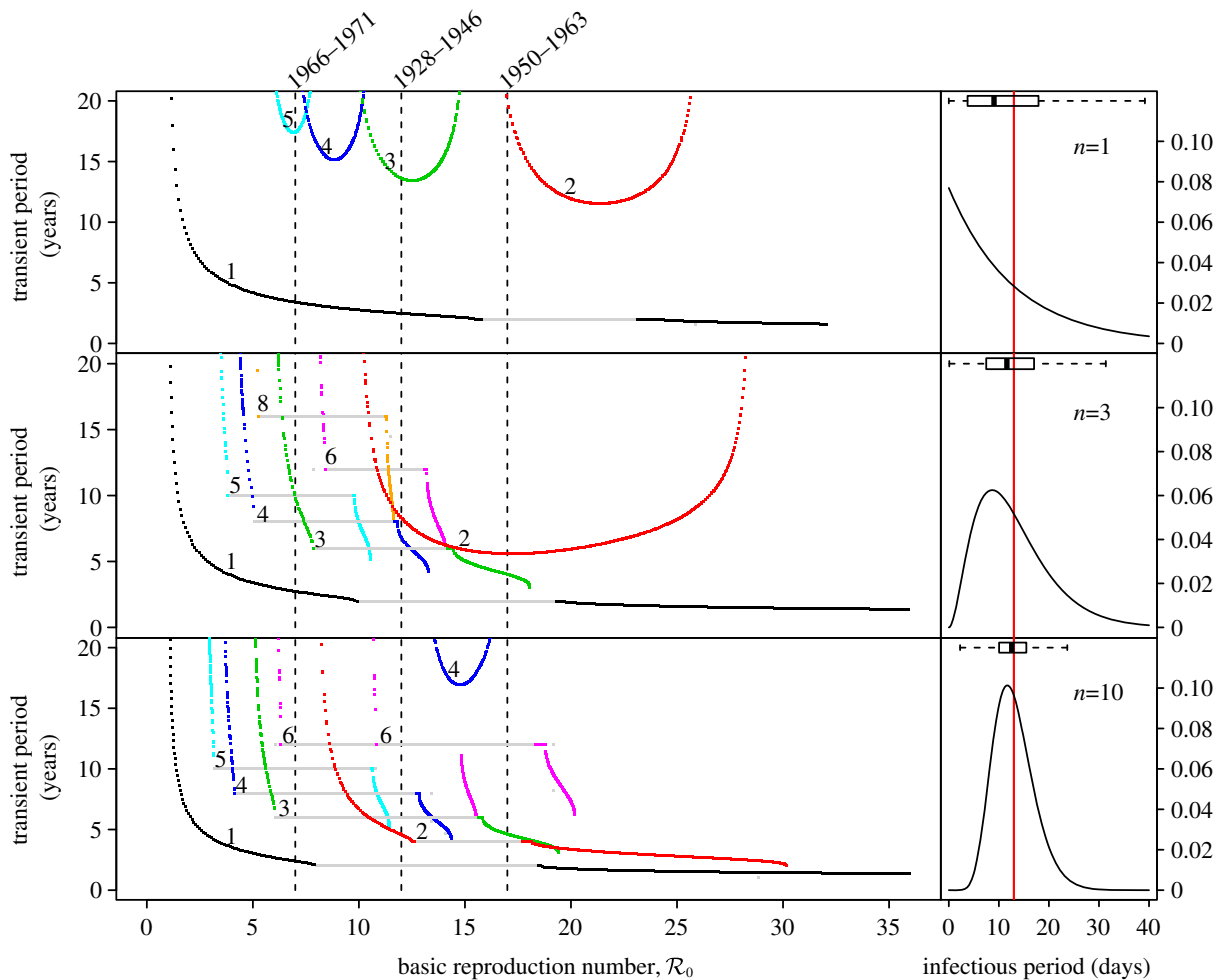
### 3.2. Asymptotic and perturbation analysis

Previous transition analyses of the New York City measles incidence time series were based on the  $SE^1I^1R$  model with mean latent and infectious periods  $\tau_E = 8$  days and  $\tau_I = 5$  days, respectively [13,15]. Given data from which the full latent and infectious period distributions can be estimated (rather than just their means), it would be sensible to fit Erlang distributions to the actual stage duration distributions and begin the transition analysis from the corresponding

$SE^mI^nR$  model. For example, Wearing & Rohani [9] used measles case data from Gloucestershire, UK, for the period 1947–1951 [53] to estimate  $\tau_E = 8$  days with the shape parameter  $m \approx 20$  and  $\tau_I = 5$  days with the shape parameter  $n \approx 20$ . Even in situations in which only the means of the stage duration distributions can be estimated, an  $SE^mI^nR$  model (with  $m > 1$  and  $n > 1$ ) is likely to be a more accurate representation of reality than an  $SE^1I^1R$  model. So, for example, Keeling & Grenfell [8] considered an  $SE^mI^nR$  model with  $m = 8$  and  $n = 5$ , i.e. one day on average in each latent and infectious substage, as a reasonable improvement of the  $SE^1I^1R$  model.

Our primary question, however, is how the predictions of transition analysis vary as a function of stage duration distribution and whether the previous transition analyses based on





**Figure 5.** Transient dynamics of the measles  $SI^nR$  model for  $n = 1, 3$  and  $10$  as a function of  $\mathcal{R}_0$ . This figure complements figure 4. Each panel shows the transient periods associated with the periodic attractors. Transient period curves are labelled according to the period of the attractor that they reach asymptotically (and in the online version are coloured correspondingly); thus the curve labelled 1 shows the transient period associated with the annual attractor shown in figure 4. The light grey lines show ranges of  $\mathcal{R}_0$  where the associated periodic solutions exist but are unstable (i.e. are repellers rather than attractors). Dashed vertical lines correspond to values of  $\mathcal{R}_{0,\text{eff}} = 7, 12$  and  $17$ . As in figure 4, the right panels show the associated infectious period distribution. (Online version in colour.)

the  $SE^{11}R$  model have led us to correct or incorrect inferences. We therefore consider the full range of Erlang distributions for the latent and infectious periods and study the  $SE^{m1}R$  model with  $1 \leq m \leq \infty$  and  $1 \leq n \leq \infty$ . Note that we chose the *mean* latent and infectious periods to be fixed ( $1/\sigma = 8$  days;  $1/\gamma = 5$  days). Because our general goal is to evaluate the robustness of dynamical inferences to model structure, we begin by analysing the simpler  $SI^nR$  model with  $1 \leq n \leq \infty$ .

### 3.2.1. Predictions of the $SI^nR$ model

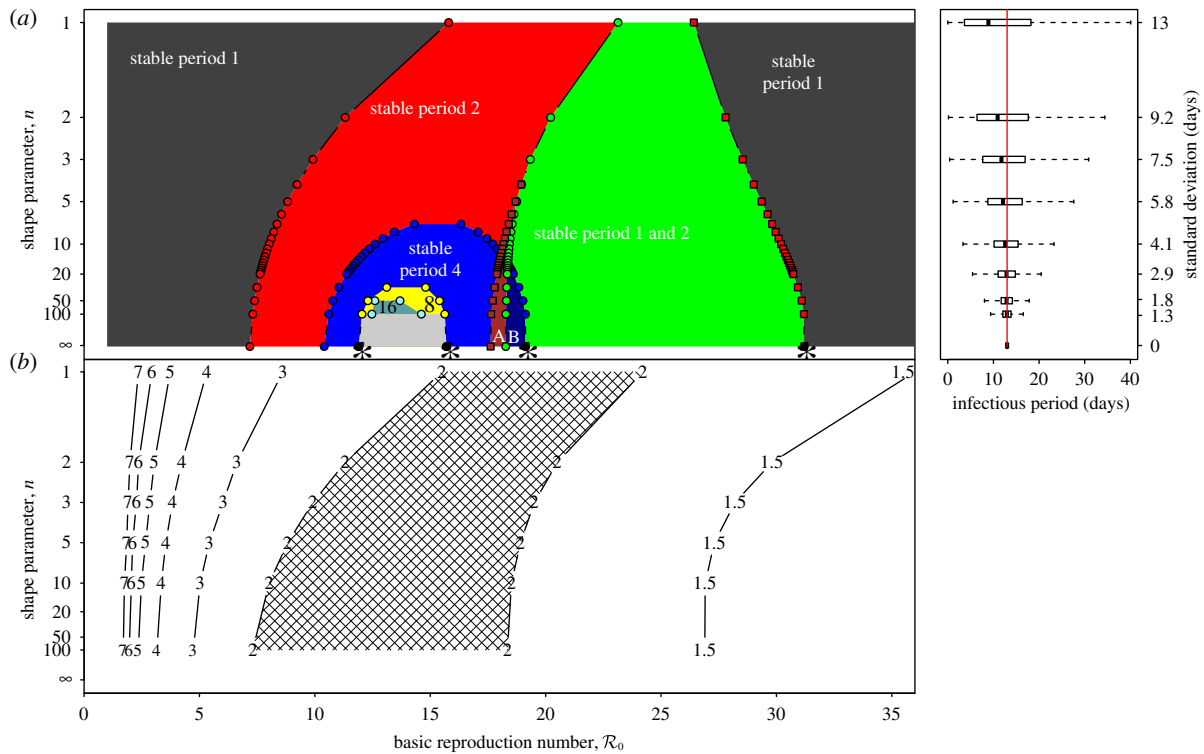
#### 3.2.1.1. Asymptotic analysis

Figure 4 shows a sequence of  $SI^nR$  bifurcation diagrams for various values of the shape parameter ( $n = 1, 3, 10, \infty$ ) together with the corresponding distributions of the infectious period (each with a mean of 13 days). Stable branches are shown as heavy curves, whereas unstable branches are shown as light curves (in the online version, stable branches of different periods are shown in different colours). The case  $n = 1$  is identical to figure 2a. As  $n$  increases from 1 to  $\infty$ , each of the branches undergoes further bifurcations. Chaotic attractors (superimposed in light grey) are evident for  $n = 10$  and dominate for a substantial range of  $\mathcal{R}_0$  for  $n = \infty$ .

The vertical dashed dark grey line at  $\mathcal{R}_0 = 17$  in figure 4 corresponds to the estimated basic reproduction number for the year  $t_0 = 1960$ . The effective reproduction number is also estimated to be 17 throughout the 13 year period  $t = 1950-1963$ , because the birth rate did not change appreciably during this time and measles vaccine was not yet invented. The other two vertical dashed grey lines at  $\mathcal{R}_0 = 7$  and  $\mathcal{R}_0 = 12$  correspond, respectively, to the estimated effective reproduction number during the periods  $t = 1928-1946$  and  $t = 1966-1971$ , as computed from equations (2.4) and (3.1).

The bifurcation tree of the standard  $SI^1R$  model ( $n = 1$ ) shows a biennial cycle for  $\mathcal{R}_0 = 17$ , coexistence of annual and triennial cycles for  $\mathcal{R}_0 = 12$ , and coexistence of annual and 4- and 5-year cycles for  $\mathcal{R}_0 = 7$ . Hence, the model correctly predicts the biennial pattern observed from 1950 to 1963 in New York City, but appears at first sight to predict incorrectly that there are multiple coexisting non-annual cycles at other times.

However, in the ranges of  $\mathcal{R}_0$  for which multiple attractors coexist, and in particular for  $\mathcal{R}_0 = 12$  and  $\mathcal{R}_0 = 7$ , stochastic simulations spend almost all of their time in the basin of the annual attractor [15]. Thus, the resonant period of 1 year observed in New York City from 1928 to 1946 and from 1966 to 1971 is also consistent with the  $SI^1R$  model.



**Figure 6.** Dynamical structure of the  $SI^nR$  model with a mean infectious period  $1/\gamma = 13$  days. (a) Two-parameter ( $\mathcal{R}_0$  versus  $n$ ) bifurcation diagram corresponding to the main branch of the one-parameter bifurcation diagrams shown in figure 4. Circles represent period-doubling (flip) bifurcations while squares denote tangent (fold) bifurcations as in figure 4. Regions are labelled according to the asymptotic dynamics on the main branch; the labels indicate which attractors are detected in each region of the  $(\mathcal{R}_0, n)$  plane: 'stable period 1' (a single annual attractor), 'stable period 2' (a single biennial attractor), 'stable period 4' (a single 4-year attractor), '8' (a single 8-year attractor), '16' (a single 16-year attractor), 'stable period 1 and 2' (coexistence of annual and biennial attractors), 'A' (coexistence of two distinct biennial attractors or coexistence of biennial and 4-year attractors) and 'B' (coexistence of annual and 4-year attractors). In the unlabelled light grey region below the region marked '16', there are cascades of further period doublings that appear to end in chaos as  $n \rightarrow \infty$ . The stars indicate bifurcation points that were estimated by extrapolation to  $n = \infty$  rather than by direct calculations based on the fixed-delay model (see the electronic supplementary material, equation S12). (b) Contours of constant transient period (associated with the annual cycle) in the  $(\mathcal{R}_0, n)$  plane (cf. black curves in figure 2b and in each panel of figure 5). In the hatched region, the transient period is phase-locked at precisely 2 years [37], whereas the transient period changes smoothly between the other contours. (Online version in colour.)

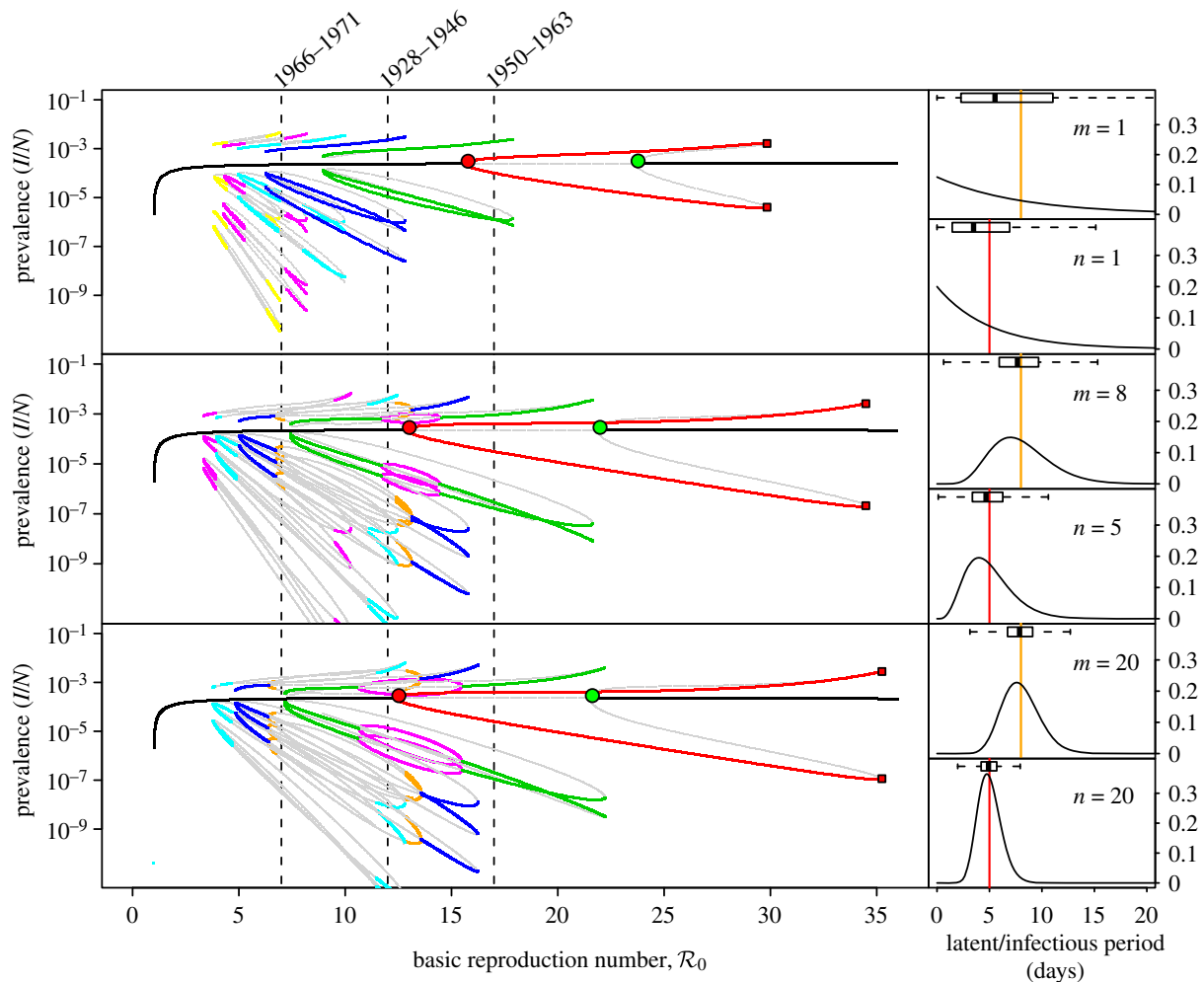
Because of the series of bifurcations that occur rapidly as  $n$  is increased, the  $SI^nR$  model for any  $n > 1$  exhibits more complex dynamics than the  $SI^1R$  model and is harder to reconcile with the observed transitions in New York City measles. More often than the  $SI^1R$  model, the  $SI^nR$  model with  $n > 1$  has coexisting long-period stable cycles that are not observed in practice. As with the  $SI^1R$  model, stochastic simulations can be expected to remain primarily in the vicinity of the 'primary' attractor, but unlike the  $SI^1R$  model, the primary attractor of the  $SI^nR$  with  $n > 1$  often predicts the wrong resonant period for New York City measles. For example, for  $n = 10$ , the dominant attractor for  $\mathcal{R}_0 = 17$  has a period of 4 years (not 2 years), and the dominant attractor for  $\mathcal{R}_0 = 12$  has period two (not one). In the presence of noise, the 4-year cycle may be difficult to distinguish from a 2-year cycle, but the predicted 2-year cycle for  $\mathcal{R}_0 = 12$  is nothing like the measles data it ought to explain.

### 3.2.1.2. Perturbation analysis

Just as perturbing an orbit away from a stable equilibrium can induce transient, damped oscillations onto the equilibrium, perturbing an orbit away from a periodic attractor can induce transient, damped oscillations onto the stable cycle. Although more cumbersome to calculate for a non-equilibrium attractor [15], transient orbits in the vicinity of a periodic attractor have a well-defined characteristic period of oscillation.

Figure 5 summarizes the transient dynamics of the  $SI^nR$  models for  $n = 1, 3$  and  $10$ . For each periodic attractor, the non-resonant period, i.e. the period of damped oscillations onto the attractor, is plotted on the  $y$ -axis as a function of  $\mathcal{R}_0$ . The curves are labelled according to the period of the corresponding attractors in figure 4. Light grey lines are used in ranges of  $\mathcal{R}_0$  where the corresponding periodic orbits are unstable; in these regions, the model displays phase-locked transient dynamics at the indicated period (i.e. the transient period is fixed and is the same as the period of the stable attractor), which is a prerequisite for a period-doubling bifurcation [37].

In the case of the  $SI^1R$  model, the non-resonant periods associated with all the non-annual attractors are too long to be observable in the New York City measles time series. The non-resonant period associated with the annual attractor does agree well with the wavelet spectrum shown in figure 3. For the  $SI^nR$  models with  $n > 1$ , the non-resonant periods associated with multi-year attractors are shorter and often should be observable in principle. For example, for  $\mathcal{R}_{0,\text{eff}} = 12$  the  $SI^{10}R$  model ( $n = 10$ ) predicts a transient period of 4.5 years. However, it is not observed in the incidence power spectra (figure 3). The lack of any indication of non-resonant periods associated with non-annual attractors in the wavelet spectrum for measles in New York City appears to cast further doubt on the usefulness of the  $SI^nR$  model for measles.



**Figure 7.**  $SE^mI^nR$  bifurcation diagrams as a function of  $\mathcal{R}_0$  for several values of the shape parameters of the latent and infectious period distributions. The mean stage durations are chosen to correspond to measles (mean latent period  $1/\sigma = 8$  days, mean infectious period  $1/\gamma = 5$  days). The other fixed parameters are the birth rate ( $\nu = 0.02$  per year) and the amplitude of (sinusoidal) seasonal forcing ( $\alpha = 0.08$ ). Heavy curves show attractors while light curves indicate unstable branches. In the online version, attractors of different periods are drawn in different colours. Circles represent period-doubling (flip) bifurcations while squares denote tangent (fold) bifurcations on the main branch. Dashed vertical lines highlight  $\mathcal{R}_{0,\text{eff}} = 7, 12$  and  $17$  associated with year ranges indicated in figure 3. The right-hand side panels show probability densities and box plots of the latent and infectious periods, with means highlighted by vertical lines and labelled according to the values of the shape parameters ( $m$  for the latent period distribution and  $n$  for the infectious period distribution). (Online version in colour.)

### 3.2.1.3. Summary of $SI^nR$ transition analysis

Overall, from the point of view of measles transition analysis, the  $SI^nR$  model is just as successful as the  $SE^1I^1R$  model studied previously [13,15]. However, the  $SI^nR$  model with  $n > 1$  is far less successful; as  $n$  increases the dynamical structure of the model becomes more and more complex and the predicted resonant and non-resonant periods stray further and further from the observed spectral peaks in the New York City measles time series.

Figure 6a summarizes our asymptotic analyses of the full sequence of  $SI^nR$  measles models ( $n = 1$  to  $\infty$ ) with a two-parameter  $(\mathcal{R}_0, n)$  bifurcation diagram for the *main branch* of the bifurcation tree in figure 4. The boundaries of the regions in figure 6 correspond to the major bifurcation points highlighted with circles (for flips) and squares (for saddle-nodes) in figure 4. As  $n \rightarrow \infty$  (i.e. as the infectious period distribution approaches a delta function), the main branch of the bifurcation tree undergoes a period-doubling cascade in the grey region ( $\mathcal{R}_0 \sim 12 - 15$ ). Figure 6b also describes the  $(\mathcal{R}_0, n)$  plane, but shows contours of constant non-resonant periods associated with the annual cycle on the main branch (this is the most likely non-resonant period to be observable because it is the

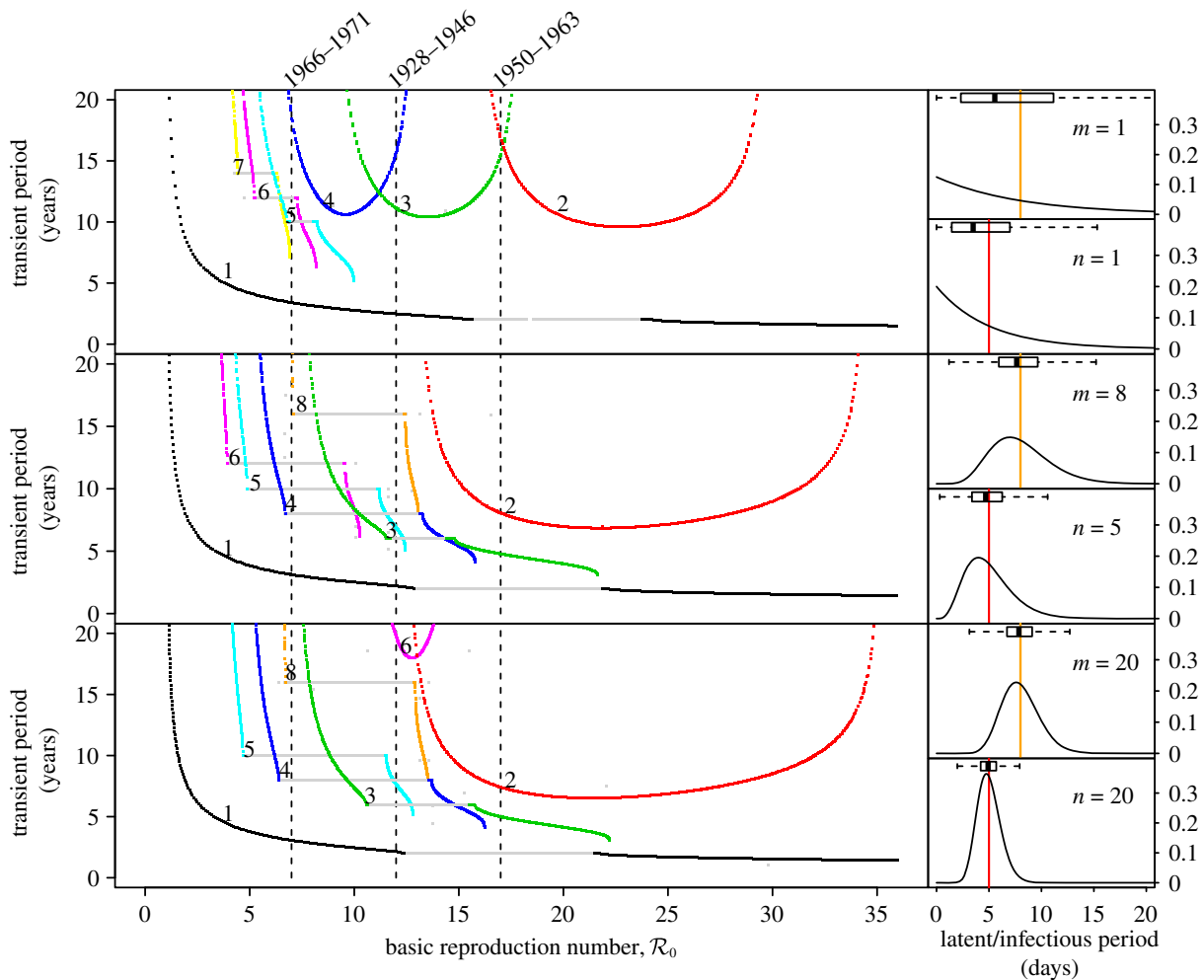
shortest; figure 5). The hatched region is characterized by phase-locked transient dynamics at a period of 2 years.

Note that because  $n$  is a discrete parameter it cannot be used as a continuation parameter in XPPAUT, hence we had to resort to separate continuation analyses for each  $n$ . The sequence of main-branch bifurcation diagrams that we constructed for the  $SI^nR$  measles model (using 24 values of  $n$  from 1 to  $\infty$ ) is shown in the electronic supplementary material, section 'Main branch of the  $SI^nR$  model'.

### 3.2.2. Predictions of the $SE^mI^nR$ model

We now apply precisely the same analyses to the more realistic  $SE^mI^nR$  models. Figures 7–9 for the  $SE^mI^nR$  models correspond to figures 4–6 for the  $SI^nR$  models.

Because we are now modelling both the latent and infectious stages directly, we can use accepted estimates for their mean durations (mean latent period  $1/\sigma = 8$  days, mean infectious period  $1/\gamma = 5$  days) [9]. In addition, we now have two shape parameters ( $m$  for the latent stage and  $n$  for the infectious stage). We examine several illustrative  $m, n$  values studied previously in the literature:  $m = 1, n = 1$  [1,13],  $m = 8, n = 5$  [8] and  $m = 20, n = 20$  [9].



**Figure 8.** Transient dynamics of the measles  $SE^mI^nR$  model for  $(m,n) = (1,1)$ ,  $(8,5)$  and  $(20,20)$  as a function of  $\mathcal{R}_0$ . Each panel shows the transient periods associated with the periodic attractors shown in figure 7. Transient period curves are labelled according to the periods of the attractors they reach asymptotically (and in the online version are coloured according to the corresponding attractor in figure 7). Light grey lines indicate regions of  $\mathcal{R}_0$  where the corresponding periodic cycles exist but are unstable. Dashed vertical lines correspond to values of  $\mathcal{R}_{0,\text{eff}} = 7, 12$  and  $17$ . As in figure 7, the right panels show the associated latent and infectious period distributions. (Online version in colour.)

Figure 7 presents asymptotic analysis of the  $SE^mI^nR$  model. The bifurcation structure of the model changes as  $m$  and  $n$  are increased, but the changes are less substantial than figure 4 shows as  $n$  is increased in the  $SI^nR$  model. Figure 8 presents the results of perturbation analysis of the  $SE^mI^nR$  model. Again, narrowing the stage duration distributions alters the transient periods, but less than figure 5 shows for the  $SI^nR$  model.

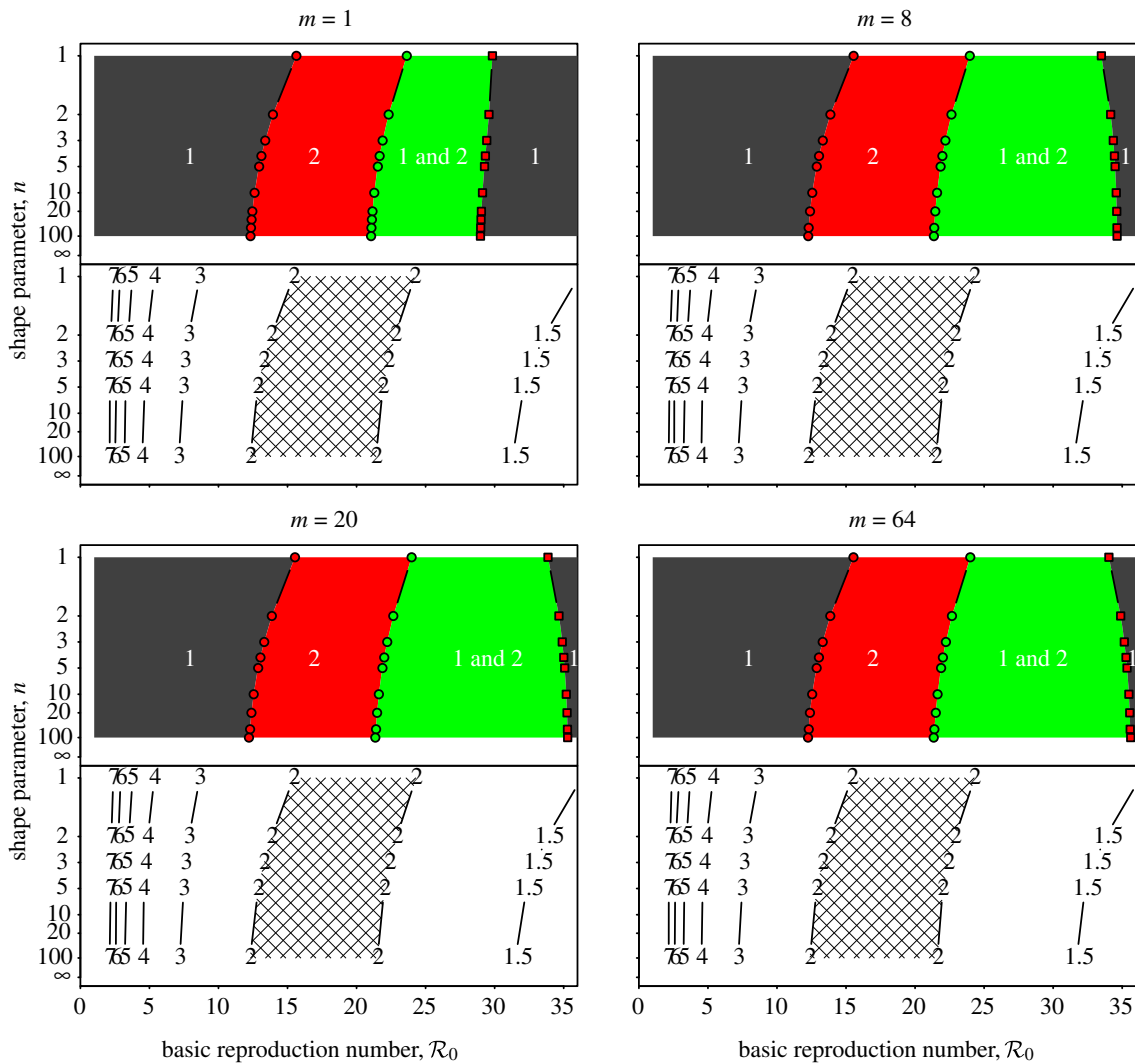
The degree of dependence of  $SE^mI^nR$  dynamics on stage duration distributions is clearest from the two-parameter bifurcation diagrams and transient-period contour plots shown in figure 9, which should be compared with figure 6 for the  $SI^nR$  model. Regardless of the shapes of the stage duration distributions, the predicted resonant and non-resonant periods are very similar. Regardless of  $m$  and  $n$ , for  $\mathcal{R}_0 = 17$ , we predict a resonant period of 2 years and an unobservably long non-resonant period (more than 7 years), for  $\mathcal{R}_0 = 12$  we predict a 1-year resonant period and a 2–3 year non-resonant period, and for  $\mathcal{R}_0 = 7$  we predict a 1-year resonant and 3–4 year non-resonant period. Consequently, transition analysis based on any of these  $SE^mI^nR$  models is consistent with the New York City measles time series and wavelet spectrum (figure 3) as well as for the other measles time series considered previously [13–15].

We are led to conclude that transition analysis is robust to the shapes of the distributions of the latent and infectious periods (provided we include both).

#### 4. The role of the generation time distribution in the dynamics of the $SI^nR$ and $SE^mI^nR$ models

It is surprising that narrowing the infectious period distribution in the  $SI^nR$  model (apparently making it more realistic) makes the model worse as a predictor of dynamical transitions (figure 6). Because the effect of narrowing the shapes of the latent and infectious period distributions in the  $SE^mI^nR$  is much smaller (figure 9), it is tempting to infer that the inclusion of a latent stage is essential for producing a robust model of the population dynamics of an infection that really does have a significant latent period. In fact, in this section, we identify the key factor that changes the structure of the  $SI^nR$  bifurcation diagram as  $n$  gets larger, and we argue ultimately that any  $SI^nR$  or  $SE^mI^nR$  model is as good as any other from the point of view of transition analysis (including the  $SI^1R$  or  $SE^1I^1R$  models) provided they are parametrized appropriately.

When using an SIR rather than SEIR model, we chose the mean infectious period to be 13 days, the sum of the actual mean latent (8 days) and mean infectious (5 days) periods.



**Figure 9.** Two-parameter bifurcation diagrams and transient-period contour plots for the measles  $SE^m I^R$  model (mean latent period  $1/\sigma = 8$  days, mean infectious period  $1/\gamma = 5$  days). Each panel corresponds to different values of the shape parameter ( $m$ ) of the latent period distribution. Regions are labelled according to the asymptotic dynamics on the main branch: '1' (single annual attractor), '2' (single biennial attractor), '1 and 2' (coexistence of annual and biennial attractors). Other annotations are as in figure 6. (Online version in colour.)

Our motivation was that it is well known that the dynamics of the unforced  $SI^1 R$  model is almost identical to that of the unforced  $SE^1 I^1 R$  model if this association is made. In particular, the period of damped oscillations about the equilibrium is then identical in the  $SI^1 R$  and  $SE^1 I^1 R$  models [1, p. 668].

It is instructive to note that the mean disease *generation time*<sup>1</sup> in the  $SE^1 I^1 R$  model is equal to the sum of the mean latent and infectious periods. So, the association we have made between the mean infectious period in the  $SI^1 R$  model and the sum of the mean latent and infectious periods in the  $SE^1 I^1 R$  model amounts to making sure both models have the same mean generation time. But for more general  $SE^m I^R$  models, the mean generation time is *not* equal to the sum of the mean latent and infectious periods. Indeed, the mean generation time in an  $SE^m I^R$  model is [55, eqn. 5.9]

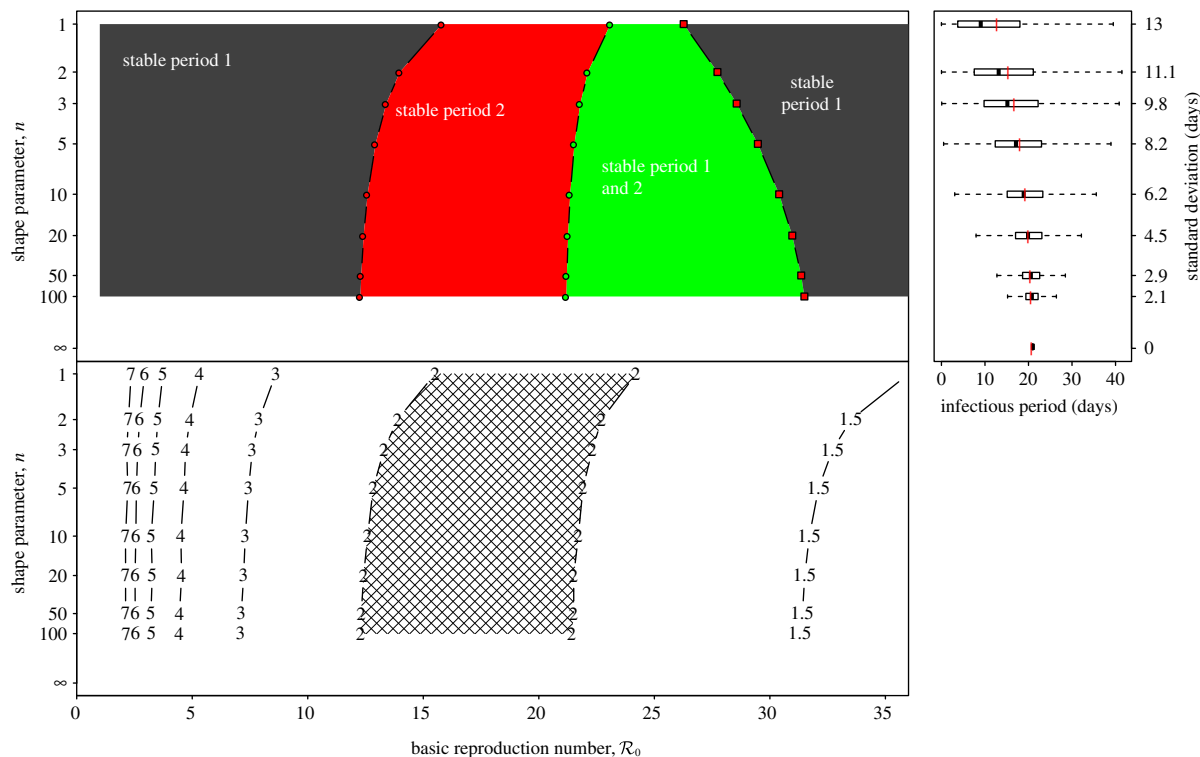
$$T_{\text{gen}} = \frac{1}{\sigma} + \left( \frac{n+1}{2n} \right) \frac{1}{\gamma}. \quad (4.1)$$

From formula (4.1), we see that the mean generation time does not depend on the shape of the latent period distribution (only its mean  $1/\sigma$ ), but decreases as the infectious period distribution gets narrower (i.e. as  $n$  increases) if the

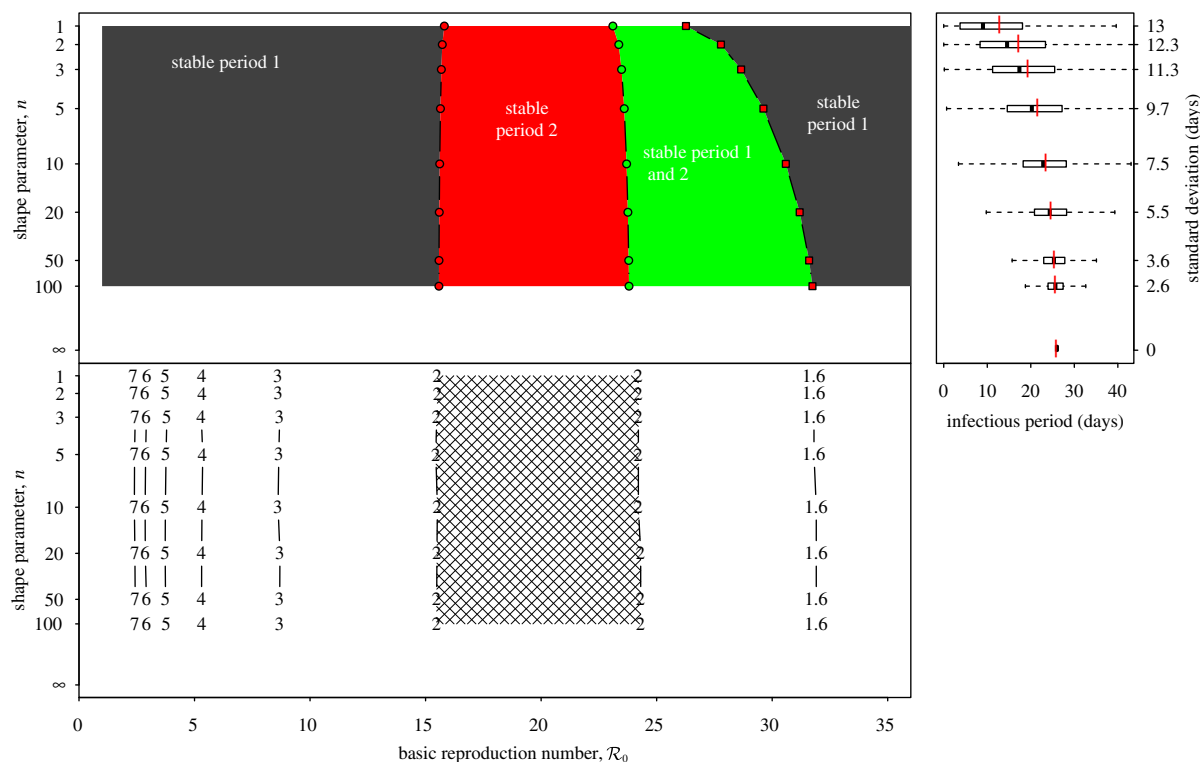
mean infectious period is kept fixed. If the mean generation time is the key factor affecting the dynamics of the  $SE^m I^R$  model then we can now easily see why figure 6 shows so much more variation than figure 9: the mean generation time  $T_{\text{gen}}$  decreases from 13 to 6.5 days as  $n$  increases from 1 to  $\infty$  in the  $SI^1 R$  model ( $1/\sigma = 0$ ,  $1/\gamma = 13$  days), whereas  $T_{\text{gen}}$  decreases only from 13 days to 10.5 days as  $n$  increases from 1 to  $\infty$  in the  $SE^m I^R$  model ( $1/\sigma = 8$  days for any value of  $m$ ,  $1/\gamma = 5$  days).

Figure 10 shows another version of the two-parameter ( $\mathcal{R}_0$  versus  $n$ ) bifurcation diagram for the  $SI^1 R$  model. Rather than fixing the mean infectious period as in figure 6, for each  $n$  we set the mean generation time to be the same as that in the  $SE^m I^R$  model with the same value of  $n$ . The result in figure 10 is now negligibly different from each of the panels of figure 9 (some details are also discussed in the electronic supplementary material, section 'Invariance of the period-doubling bifurcation point').

Finally, in figure 11, we show yet another version of the  $\mathcal{R}_0$  versus  $n$  bifurcation diagram for the  $SI^1 R$  model, this time keeping the mean generation time fixed at 13 days for all values of  $n$  (in contrast to figure 10, where the mean generation time for each  $SI^1 R$  model was chosen to be the same as in the



**Figure 10.** Two-parameter bifurcation diagram and transient-period contour plot for the measles  $SI^nR$  model with the mean generation time chosen to be the same as in the  $SE^mI^nR$  model for each value of the shape parameter of the infectious period distribution,  $n$ . Annotation is as in figure 6. Note the similarity between this figure and the  $SE^mI^nR$  model diagrams in figure 9. (Online version in colour.)



**Figure 11.** Two-parameter bifurcation diagram and transient-period contour plot for the measles  $SI^nR$  model with fixed mean generation time,  $T_{gen} = 13$  days. Annotation is as in figure 6. Note that the main period-doubling bifurcation point from annual to biennial cycle occurs for approximately the same value of  $\mathcal{R}_0$ , regardless of the shape ( $n$ ) of the infectious period distribution. The bottom panel shows that the transient period associated with the annual cycle is almost exactly the same for all values of  $n$ , i.e. this transient period is almost independent of the shape of the generation time distribution. (Online version in colour.)

$SE^mI^nR$  model with the same value of  $n$ ). Figure 11 makes clear that from the point of view of transition analysis—and to a large extent more generally for understanding the dynamics of

$SE^mI^nR$  models—the key parameter that needs to be estimated is the mean generation time, not the mean latent or mean infectious period themselves and certainly not the shapes of these

distributions. For a given mean generation time, it makes little difference which  $SI^R$  or  $SE^{III}R$  model we use, so we might as well work with the simplest, the  $SI^R$  model.

## 5. Discussion

We set out to determine whether the results of previous 'transition analyses' of recurrent epidemic patterns of childhood diseases [3,13,15] were robust to the assumed shapes of the latent and infectious period distributions (which were taken to be exponential in previous work). We focused on measles and undertook a systematic analysis of the sequence of  $SI^R$  and  $SE^{III}R$  models for measles, and concluded that for a given mean generation time, transition analyses based on any  $SI^R$  or  $SE^{III}R$  model will lead to the same predictions for measles. Consequently, transition analyses of measles dynamics can be safely conducted using the very simplest  $SI^R$  model. It is important to emphasize, however, that the mean generation time must be estimated correctly for this to work; in particular, it is not true that the real mean generation time is the sum of the mean latent and infectious periods.

The key graph that establishes that  $SI^R$  dynamics are nearly invariant for measles, if the mean generation time is fixed, is figure 11 (where the mean generation time is set to 13 days). In future work, we will construct the equivalent graph for a sequence of mean generation times that covers the range of typical recurrent infectious diseases, in order to determine whether transition analyses of other diseases

can also be safely conducted with the simple  $SI^R$  model. There is also considerable scope for analytical developments that complement our numerical analysis and build on previous analytical work associated with the role of the generation time distribution [56–59].

Consistent with previous work [6,7,10], we found that if we fix the mean infectious period (rather than the mean generation time) then narrowing the infectious period distribution (which reduces the mean generation time) leads to more complex dynamics. Previous work has also investigated the stochastic dynamics of  $SI^R$  and  $SE^{III}R$  models and examined characteristics such as the critical community size for disease persistence [6,12]. In future work, we will re-examine inferences concerning the stochastic dynamics of these models in light of the now-evident importance of the mean generation time for their deterministic dynamics.

We thank Jonathan Dushoff and the other members of the Mathematical Biology Group at McMaster University for helpful comments and discussions. We were supported by the Natural Sciences and Engineering Research Council of Canada (O.K. by an NSERC Postgraduate Scholarship and D.J.D.E. by an NSERC Discovery grant). The data used in this paper can be downloaded from the International Infectious Disease Data Archive (<http://iidda.mcmaster.ca>).

## Endnote

<sup>1</sup>The generation time is also called the generation interval, the serial interval or the case-to-case interval. It is the time from initial infection of a primary case to initial infection of a secondary case [54].

## References

- Anderson RM, May RM. 1991 *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford University Press.
- Hethcote HW. 2000 The mathematics of infectious diseases. *SIAM Rev.* **42**, 599–653. (doi:10.1137/S0036144500371907)
- Earn DJD. 2009 Mathematical epidemiology of infectious diseases. In *Mathematical biology* vol. 14, (eds MA Lewis, MAJ Chaplain, JP Keener, PK Maini), pp. 151–186. IAS/Park City Mathematics Series. Providence, RI: American Mathematical Society.
- Kermack WO, McKendrick AG. 1927 Contributions to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A* **115**, 700–721. (doi:10.1098/rspa.1927.0118)
- Brauer F. 2008 Compartmental models in epidemiology. In *Mathematical epidemiology*, vol. 1945 of *Lecture notes in mathematics*, pp. 19–79. Berlin, Germany: Springer.
- Lloyd AL. 2001 Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proc. R. Soc. Lond. B* **268**, 985–993. (doi:10.1098/rspb.2001.1599)
- Lloyd AL. 2001 Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theor. Popul. Biol.* **60**, 59–71. (doi:10.1006/tpbi.2001.1525)
- Keeling MJ, Grenfell BT. 2001 Understanding the persistence of measles: reconciling theory, simulation and observation. *Proc. R. Soc. Lond. B* **269**, 335–343. (doi:10.1098/rspb.2001.1898)
- Wearing HJ, Rohani P, Keeling MJ. 2005 Appropriate models for the management of infectious diseases. *PLoS Med.* **2**, 621–627. (doi:10.1371/journal.pmed.0020174)
- Nguyen HTH, Rohani P. 2008 Noise, nonlinearity and seasonality: the epidemics of whooping cough revisited. *J. R. Soc. Interface* **5**, 403–413. (doi:10.1098/rsif.2007.1168)
- Black AJ, McKane AJ. 2010 Stochasticity in staged models of epidemics: quantifying the dynamics of whooping cough. *J. R. Soc. Interface* **7**, 1219–1227. (doi:10.1098/rsif.2009.0514)
- Conlan n, Rohani P, Lloyd AL, Keeling M, Grenfell BT. 2010 Resolving the impact of waiting time distributions on the persistence of measles. *J. R. Soc. Interface* **7**, 623–640. (doi:10.1098/rsif.2009.0284)
- Earn DJD, Rohani P, Bolker BM, Grenfell BT. 2000 A simple model for complex dynamical transitions in epidemics. *Science* **287**, 667–670. (doi:10.1126/science.287.5453.667)
- Bauch CT, Earn DJD. 2003 Interepidemic intervals in forced and unforced SEIR models. *Fields Inst. Commun.* **36**, 33–44.
- Bauch CT, Earn DJD. 2003 Transients and attractors in epidemics. *Proc. R. Soc. Lond. B* **207**, 1573–1578. (doi:10.1098/rspb.2003.2410)
- Bailey NTJ. 1956 On estimating of latent and infectious periods of measles. I. Families with two susceptibles only. *Biometrika* **43**, 15–22.
- Bailey NTJ. 1956 On estimating of latent and infectious periods of measles. II. Families with three or more susceptibles. *Biometrika* **43**, 322–331.
- Gough KJ. 1977 The estimation of latent and infectious periods. *Biometrika* **64**, 559–565. (doi:10.1093/biomet/64.3.559)
- Nishiura H, Eichner M. 2007 Infectiousness of smallpox relative to disease age: estimates based on transmission network and incubation period. *Epidemiol. Infect.* **135**, 1145–1150. (doi:10.1017/S0950268806007618)
- Nishiura H. 2007 Early efforts in modeling the incubation period of infectious diseases with an acute course of illness. *Emerg. Themes Epidemiol.* **4**, 2. (doi:10.1186/1742-7622-4-2)
- Eichner M, Dietz K. 2003 Transmission potential of smallpox: estimates based on detailed data from an outbreak. *Am. J. Epidemiol.* **158**, 110–117. (doi:10.1093/aje/kwg103)
- Bailey N. 1964 Some stochastic models for small epidemics in large populations. *Appl. Stat.* **13**, 9–19.
- Anderson D, Watson R. 1980 On the spread of a disease with gamma distributed latent and infectious periods. *Biometrika* **67**, 191–198. (doi:10.1093/biomet/67.1.191)

24. Ma J, Earn DJD. 2006 Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bull. Math. Biol.* **68**, 679–702. (doi:10.1007/s11538-005-9047-7)
25. London WP, Yorke JA. 1973 Recurrent outbreaks of measles, chickenpox and mumps. I. Seasonal variation in contact rates. *Am. J. Epidemiol.* **98**, 453–468.
26. Bacaër N. 2007 Approximation of the basic reproduction number  $R_0$  for vector-borne diseases with a periodic vector population. *Bull. Math. Biol.* **69**, 1067–1091. (doi:10.1007/s11538-006-9166-9)
27. Bacaër N, Ait Dads EH. 2011 Genealogy with seasonality, the basic reproduction number, and the influenza pandemic. *J. Math. Biol.* **62**, 741–762. (doi:10.1007/s00285-010-0354-8)
28. Ma J, Ma Z. 2006 Epidemic threshold conditions for seasonally forced SEIR models. *Math. Biosci. Eng.* **3**, 161–172.
29. Hethcote HW, Tudor DW. 1980 Integral equation models for endemic infectious diseases. *J. Math. Biol.* **9**, 37–47. (doi:10.1007/BF00276034)
30. Feng Z, Thieme HR. 2000 Endemic models with arbitrarily distributed periods of infection I: fundamental properties of the model. *SIAM J. Appl. Math.* **61**, 803–833. (doi:10.1137/S0036139998347834)
31. Feng Z, Thieme HR. 2000 Endemic models with arbitrarily distributed periods of infection II: fundamental properties of the model. *SIAM J. Appl. Math.* **61**, 983–1012.
32. Zhang F, Li Z, Zhang F. 2008 Global stability of an SIR epidemic model with constant infectious period. *Appl. Math. Comput.* **199**, 285–291. (doi:10.1016/j.amc.2007.09.053)
33. Grossman Z. 1980 Oscillatory phenomenon in a model of infectious diseases. *Theor. Popul. Biol.* **18**, 204–243. (doi:10.1016/0040-5809(80)90050-7)
34. Yorke JA, London WP. 1973 Recurrent outbreaks of measles, chickenpox and mumps. II. Systematic differences in contact rates and stochastic effects. *Am. J. Epidemiol.* **98**, 469–482.
35. Olsen LF, Schaffer WM. 1990 Chaos versus noisy periodicity: alternative hypotheses for childhood epidemics. *Science* **249**, 499–504. (doi:10.1126/science.2382131)
36. Bolker BM, Grenfell BT. 1993 Chaos and biological complexity in measles dynamics. *Proc. R. Soc. Lond. B* **251**, 75–81. (doi:10.1098/rspb.1993.0011)
37. He D, Earn DJD. 2007 Epidemiological effects of seasonal oscillations in birth rates. *Theor. Popul. Biol.* **72**, 274–291. (doi:10.1016/j.tpb.2007.04.004)
38. Heesterbeek JAP, Metz JAJ. 1996 The saturating contact rate in epidemic models. In *Models for infectious human diseases* (eds V Isham, G Medley), pp. 308–310. Cambridge, UK: Cambridge University Press.
39. Grenfell BT, Bjornstad ON, Kappey J. 2001 Travelling waves and spatial hierarchies in measles epidemics. *Nature* **414**, 716–723. (doi:10.1038/414716a)
40. Bauch CT. 2008 The role of mathematical models in explaining recurrent outbreaks of infectious childhood diseases. In *Lecture notes in mathematical epidemiology*, vol. 1945 of *Lecture notes in mathematics* (eds F Brauer, P van den Driessche, J Wu), pp. 297–319. Berlin, Germany: Springer.
41. Ermentrout B. 2002 *Simulating, analyzing, and animating dynamical systems: a guide to XPPAUT for researchers and students*. Philadelphia, PA: Society for Industrial and Applied Mathematics.
42. Bartlett MS. 1957 Measles periodicity and community size. *J. R. Stat. Soc. A* **120**, 48–70. (doi:10.2307/2342553)
43. Alonso D, McKane AJ, Pascual M. 2007 Stochastic amplification in epidemics. *J. R. Soc. Interface* **4**, 575–582. (doi:10.1098/rsif.2006.0192)
44. Lima M. 2009 A link between the North Atlantic oscillation and measles dynamics during the vaccination period in England and Wales. *Ecol. Lett.* **12**, 302–314. (doi:10.1111/j.1461-0248.2009.01289.x)
45. Black AJ, McKane AJ. 2010 Stochastic amplification in an epidemic model with seasonal forcing. *J. Theor. Biol.* **267**, 85–94. (doi:10.1016/j.jtbi.2010.08.014)
46. Schwartz IB, Smith HL. 1983 Infinite subharmonic bifurcation in SEIR epidemic model. *J. Math. Biol.* **18**, 233–253. (doi:10.1007/BF00276090)
47. Rand DA, Wilson HB. 1991 Chaotic stochasticity: a ubiquitous source of unpredictability in epidemics. *Proc. R. Soc. Lond. B* **246**, 179–184. (doi:10.1098/rspb.1991.0142)
48. Cazelles B, Chavez M, Berteaux D, Menard F, Vik JO, Jenouvrier S, Stenseth NC. 2008 Wavelet analysis of ecological time series. *Oecologia* **156**, 287–304. (doi:10.1007/s00442-008-0993-2)
49. Torrence C, Compo GP. 1998 A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* **79**, 61–78. (doi:10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2)
50. Cazelles B, Chavez M, de Magny GC, Guegan JF, Hales S. 2007 Time-dependent spectral analysis of epidemiological time-series with wavelets. *J. R. Soc. Interface* **4**, 625–636. (doi:10.1098/rsif.2007.0212)
51. Measles, mumps, rubella. 2010 Galveston, TX: National Network for Immunization Information. See <http://www.immunizationinfo.org/vaccines/measles>.
52. Vital statistics of the United States. 2011 Hyattsville, MD: Center for Disease Control and Prevention. National Center for Health Statistics. See <http://www.cdc.gov/nchs/products/vsus.htm>.
53. Hope-Simpson RE. 1952 Infectiousness of communicable diseases in the household (measles, chickenpox, and mumps). *Lancet* **2**, 549–554.
54. Fine PEM. 2003 The interval between successive cases of an infectious disease. *Am. J. Epidemiol.* **158**, 1039–1047. (doi:10.1093/aje/kwg251)
55. Svensson A. 2007 A note on generation times in epidemic models. *Math. Biosci.* **208**, 300–311. (doi:10.1016/j.mbs.2006.10.010)
56. Demetrius L. 1977 Measures of fitness and demographic stability. *Proc. Natl Acad. Sci. USA* **74**, 384–386.
57. Wallinga J, Lipsitch M. 2007 How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. B* **274**, 599–604. (doi:10.3410/f.717988268.793472595)
58. Roberts MG, Heesterbeek JAP. 2007 Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection. *J. Math. Biol.* **55**, 803–816. (doi:10.1007/s00285-007-0112-8)
59. Roberts MG, Nishiura H. 2011 Early estimation of the reproduction number in the presence of imported cases: pandemic influenza H1N1-2009 in New Zealand. *PLoS ONE* **6**, e17835. (doi:10.1371/journal.pone.0017835.)



*Electronic Supplementary Materials for*  
“Effects of the infectious period distribution  
on predicted transitions in childhood disease  
dynamics”

Olga Krylova and David J.D. Earn

*Department of Mathematics and Statistics, McMaster University,  
Hamilton, Ontario, Canada L8S 4K1*

April 14, 2013

# Contents

<b>1</b>	<b>Models</b>	<b>S2</b>
<b>2</b>	<b>Vaccination level calculations</b>	<b>S9</b>
<b>3</b>	<b>Invariance of the period doubling bifurcation point</b>	<b>S11</b>
<b>4</b>	<b>Main branch of the <math>SI^nR</math> model with fixed mean infectious period, <math>\frac{1}{\gamma} = 13</math> days</b>	<b>S13</b>

# 1 Models

## SI<sup>1</sup>R and SE<sup>1</sup>I<sup>1</sup>R

Assuming the population is large and homogeneously mixed, the (unforced) SI<sup>1</sup>R model can be cast as a simple system of nonlinear ordinary differential equations [1, 2]:

$$\frac{dS}{dt} = \nu - \beta SI - \mu S, \quad (\text{S2a})$$

$$\frac{dI}{dt} = \beta SI - \gamma I - \mu I, \quad (\text{S2b})$$

$$\frac{dR}{dt} = \gamma I - \mu R. \quad (\text{S2c})$$

Here,  $S$ ,  $I$  and  $R$  are the numbers of susceptible, infectious, and recovered (immune) individuals in the population.  $\mu$ ,  $\beta$  and  $\gamma$  are the rates of *per capita* death, transmission and recovery, respectively.  $\mu$  quantifies death from “natural causes” (disease-induced mortality is assumed to be negligible).  $\nu$  denotes the number of births per unit time, which is often time-dependent in practice [3]. If  $\nu = \mu N$ , where  $N = S + I + R$  is the total population size, then births balance deaths and the population size remains constant.  $\beta$  is the rate at which contacts between susceptible and infectious individuals cause new infections (per susceptible per infected), so  $\beta SI$  is the number of new infections that occur per unit time (*incidence rate*). Note that equations S2a and S2b do not depend on  $R$ . Therefore they completely specify the system dynamics and equation S2c can be ignored.

For our purposes, the birth term ( $\nu$ ) is particularly important because secular changes in this term can induce dynamical transitions [4, 5, 6, 7]. We estimate  $\nu$  based on demographic data and do not assume that it scales with population size (e.g., we do *not* assume  $\nu = \mu N$ ). Nevertheless, it is convenient to express  $\nu$  in units that are similar to those of the *per capita* death rate  $\mu$ . We therefore write  $\nu = \nu N_0$ , where  $N_0$  is the population size at a particular “anchor time”  $t_0$  (see also section 2.2).  $\nu$  represents births *per capita* at time  $t_0$ , but not at other times.

We rewrite equation S2 as

$$\frac{dS}{dt} = \nu N_0 - \beta SI - \mu S, \quad (\text{S3a})$$

$$\frac{dI}{dt} = \beta SI - \gamma I - \mu I, \quad (\text{S3b})$$

$$\frac{dR}{dt} = \gamma I - \mu R. \quad (\text{S3c})$$

In the special case in which births balance deaths ( $\nu = \mu$ ), the *basic reproduction number*,  $\mathcal{R}_0$ , for this model is well-known to be  $N_0\beta/(\gamma + \mu)$  [1]. To see how the expression for  $\mathcal{R}_0$  changes if  $\nu \neq \mu$ , consider the scaled variables

$$\tilde{S} = \frac{\mu}{\nu} S, \quad \tilde{I} = \frac{\mu}{\nu} I, \quad \tilde{R} = \frac{\mu}{\nu} R. \quad (\text{S4})$$

In these variables, Equations S3 become

$$\frac{d\tilde{S}}{dt} = \mu N_0 - \frac{\nu\beta}{\mu} \tilde{S}\tilde{I} - \mu\tilde{S}, \quad (\text{S5a})$$

$$\frac{d\tilde{I}}{dt} = \frac{\nu\beta}{\mu} \tilde{S}\tilde{I} - \gamma\tilde{I} - \mu\tilde{I}, \quad (\text{S5b})$$

$$\frac{d\tilde{R}}{dt} = \gamma\tilde{I} - \mu\tilde{R}. \quad (\text{S5c})$$

These equations are exactly equivalent to the standard SI<sup>1</sup>R model with births balancing deaths (both rates equal to  $\mu$ ), but with transmission rate  $\nu\beta/\mu$  rather than  $\beta$ . Consequently, the threshold for disease spread is

$$\mathcal{R}_0^{\text{SI}^1\text{R}} = \frac{\nu N_0}{\mu} \frac{\beta}{\gamma + \mu}. \quad (\text{S6})$$

We assume that  $\nu$  changes slowly enough that it can be regarded as constant for the purposes of defining  $\mathcal{R}_0$  at a given time.

The SI<sup>1</sup>R model can easily be extended to the SE<sup>1</sup>I<sup>1</sup>R model, which includes a *latent stage*, by replacing equation S3b with the two equations:

$$\frac{dE}{dt} = \beta SI - \sigma E - \mu E, \quad (\text{S7a})$$

$$\frac{dI}{dt} = \sigma E - \gamma I - \mu I. \quad (\text{S7b})$$

The latent period is defined to be the time from initial infection to becoming infectious. In equations S7a and S7b  $E$  represents the number of *exposed* individuals (individuals in the latent stage). For the  $SE^1I^1R$  model, the basic reproduction number is

$$\mathcal{R}_0^{SE^1I^1R} = \frac{\nu N_0}{\mu} \frac{\beta \sigma}{(\sigma + \mu)(\gamma + \mu)}. \quad (\text{S8})$$

In the standard  $SI^1R$  and  $SE^1I^1R$  formulation (equations S3 and S7), the lengths of the latent and infectious stages are exponentially distributed. To see this, suppose that during the infectious stage the only process occurring was recovery from infection. Then equation S3b would reduce to  $dI/dt = -\gamma I$ , which implies that the distribution of time spent in the infectious class (the infectious period) is exponential with mean  $1/\gamma$  (if  $I_0$  individuals are infectious at time 0 then  $I_0 e^{-\gamma t}$  are still infectious at time  $t$ ). Similarly the latent period is exponentially distributed with mean  $1/\sigma$ .

### **$SI^nR$ and $SE^mI^nR$**

Arbitrarily distributed stage durations can be included into SIR and SEIR models via integro-differential equations [8, 9, 10]. Unfortunately, the resulting dynamical systems are mathematically and computationally difficult to study. To avoid the complications involved with integro-differential equations, most research on epidemic models with non-exponentially distributed stage durations has restricted attention to a convenient class of realistic (but not arbitrary) distributions, namely Gamma distributions with integer shape parameter (also known as *Erlang distributions*) [11, 12, 13, 8]. The idea is to exploit the fact that the sum of a sequence of independent exponentially distributed random variables is Gamma distributed [14]. If we break up the infectious stage into a sequence of  $n$  substages, each exponentially distributed with mean  $1/(n\gamma)$ , then the full infectious period distribution will be the Erlang distribution with shape parameter  $n$  and scale parameter  $n\gamma$ , Erlang( $n, n\gamma$ ). The resulting  $SI^nR$  model can then be represented by a simple

system of ODEs:

$$\frac{dS}{dt} = \nu N_0 - \beta SI - \mu S, \quad (\text{S9a})$$

$$\frac{dI_1}{dt} = \beta SI - (n\gamma + \mu)I_1, \quad (\text{S9b})$$

$$\frac{dI_2}{dt} = n\gamma I_1 - (n\gamma + \mu)I_2, \quad (\text{S9c})$$

⋮

$$\frac{dI_n}{dt} = n\gamma I_{n-1} - (n\gamma + \mu)I_n, \quad (\text{S9d})$$

and the basic reproduction number is [15]

$$\mathcal{R}_0^{\text{SI}^n\text{R}} = \frac{\nu N_0}{\mu} \frac{\beta}{n\gamma + \mu} \sum_{j=0}^{n-1} \left( \frac{n\gamma}{n\gamma + \mu} \right)^j. \quad (\text{S10})$$

Note that the number of individuals in the infectious stage ( $I$ ) is the sum of all individuals currently in each infectious substage,

$$I = \sum_{j=1}^n I_j. \quad (\text{S11})$$

Division into  $n$  subclasses is purely a mathematical device and has no biological meaning.

The two extreme cases of the  $\text{SI}^n\text{R}$  model occur for  $n = 1$ , in which case the model reduces to the standard  $\text{SI}^1\text{R}$  model (equation S3), and the limit as  $n \rightarrow \infty$ , which yields a fixed infectious period of  $\tau = 1/\gamma$  (equation 1, see also figure 1), (i.e., the infectious period has a Dirac delta distribution  $\delta(t - \tau)$ : all individuals who become infectious at time  $t$  recover at exactly time  $t + \tau$ ). In this limit, the system becomes a delay differential equation, which can be seen directly as follows. Since the incidence rate at time  $t$  is  $\beta(t)S(t)I(t)$  and the probability that an individual alive at time  $t$  survives to time  $t + \tau$  is  $e^{-\mu\tau}$  [9], the recovery rate at time  $t$  is  $\beta(t - \tau)e^{-\mu\tau}S(t - \tau)I(t - \tau)$ . Thus, in the limit  $n \rightarrow \infty$ , the  $\text{SI}^n\text{R}$

model approaches the system

$$\frac{dS}{dt} = \nu N_0 - \beta(t)S(t)I(t) - \mu S(t), \quad (\text{S12a})$$

$$\frac{dI}{dt} = \beta(t)S(t)I(t) - \beta(t - \tau)e^{-\mu\tau}S(t - \tau)I(t - \tau) - \mu I(t), \quad (\text{S12b})$$

$$\frac{dR}{dt} = \beta(t - \tau)e^{-\mu\tau}S(t - \tau)I(t - \tau) - \mu R(t). \quad (\text{S12c})$$

We obtain the  $SE^mI^nR$  model (with mean latent period  $\tau_E = 1/\sigma$  and mean infectious period  $\tau_I = 1/\gamma$ ) by subdividing the exposed class into  $m$  subclasses [16, 17],

$$\frac{dS}{dt} = \nu N_0 - \beta SI - \mu S, \quad (\text{S13a})$$

$$\frac{dE_1}{dt} = \beta SI - (m\sigma + \mu)E_1, \quad (\text{S13b})$$

$$\frac{dE_2}{dt} = m\sigma E_1 - (m\sigma + \mu)E_2, \quad (\text{S13c})$$

⋮

$$\frac{dE_m}{dt} = m\sigma E_{m-1} - (m\sigma + \mu)E_m, \quad (\text{S13d})$$

$$\frac{dI_1}{dt} = m\sigma E_m - (n\gamma + \mu)I_1, \quad (\text{S13e})$$

$$\frac{dI_2}{dt} = n\gamma I_1 - (n\gamma + \mu)I_2, \quad (\text{S13f})$$

⋮

$$\frac{dI_n}{dt} = n\gamma I_{n-1} - (n\gamma + \mu)I_n. \quad (\text{S13g})$$

In the limit that  $m \rightarrow \infty$  and  $n \rightarrow \infty$  we obtain the delay differential equation [16],

$$\frac{dS}{dt} = \nu N_0 - \beta(t)S(t)I(t) - \mu S(t), \quad (\text{S14a})$$

$$\frac{dE}{dt} = \beta(t)S(t)I(t) - \beta(t - \tau_E)e^{-\mu\tau_E}S(t - \tau_E)I(t - \tau_E) - \mu E(t), \quad (\text{S14b})$$

$$\frac{dI}{dt} = \beta(t - \tau_E)e^{-\mu\tau_E}S(t - \tau_E)I(t - \tau_E) - \beta(t - \tau_I)e^{-\mu\tau_I}S(t - \tau_I)I(t - \tau_I) - \mu I(t), \quad (\text{S14c})$$

$$- \beta(t - \tau_I)e^{-\mu\tau_I}S(t - \tau_I)I(t - \tau_I) - \mu I(t), \quad (\text{S14d})$$

$$\frac{dR}{dt} = \beta(t - \tau_I)e^{-\mu\tau_I}S(t - \tau_I)I(t - \tau_I) - \mu R(t). \quad (\text{S14e})$$

For the  $\text{SE}^m\text{I}^n\text{R}$  model  $\mathcal{R}_0$  can be computed by the following formula [15]:

$$\mathcal{R}_0^{\text{SE}^m\text{I}^n\text{R}} = \frac{\nu N_0}{\mu} \left( \frac{m\sigma}{m\sigma + \mu} \right)^m \frac{\beta}{n\gamma + \mu} \sum_{j=0}^{n-1} \left( \frac{n\gamma}{n\gamma + \mu} \right)^j. \quad (\text{S15})$$

Note that the basic reproduction number is the product of the mean transmission rate and the mean duration of infectiousness. Altering the distribution of the infectious period alters the probability that an infectious host will die before transmitting. This death-delay interaction changes the mean duration of infectiousness, which results in the differences in the formulae for  $\mathcal{R}_0$  above. For diseases of short duration such as measles and smallpox, the mean host lifetime is much longer than the duration of infectiousness. Consequently,  $\mu \ll \sigma$  and  $\mu \ll \gamma$ , so  $\mathcal{R}_0$  can always be written:

$$\mathcal{R}_0 \approx \frac{\nu N_0 \beta}{\mu \gamma}. \quad (\text{S16})$$

## Seasonal forcing

Exponentially distributed disease stage durations is one unrealistic assumption used in standard SIR-type models. Another is treating the transmission rate  $\beta$  as a constant. More realistic seasonally forced models are implemented by allowing the transmission rate to vary periodically with a period of one year. The two most commonly used seasonal patterns are *sinusoidal forcing* [16, 18, 19, 20],

$$\beta(t) = \langle \beta \rangle (1 + \alpha \cos(2\pi t)), \quad (\text{S17})$$



with mean  $\langle\beta\rangle$  and amplitude  $\alpha$  ( $0 \leq \alpha \leq 1$ ) and *term-time forcing* [16, 17, 5, 7, 21, 22]

$$\beta(t) = \begin{cases} \beta_{\text{H}} & \text{school days,} \\ \beta_{\text{L}} & \text{non-school days,} \end{cases} \quad (\text{S18})$$

where  $\beta_{\text{H}} > \beta_{\text{L}}$  (the transmission rate is high when school is in session and low otherwise). Earn *et al.* [5] found that the qualitative dynamics of the term-time forced  $\text{SE}^1\text{I}^1\text{R}$  model are essentially equivalent to the dynamics of the sinusoidally forced  $\text{SE}^1\text{I}^1\text{R}$  model but with lower seasonal amplitude,  $\alpha$ . The same is true for the  $\text{SI}^1\text{R}$  model. Since our focus is on qualitative dynamics, we use sinusoidal forcing for simplicity.

## 2 Vaccination level calculations

Live measles virus vaccine was licensed in the US in 1963 [23]. A national campaign to eliminate measles was launched by the Center for Disease Control (CDC) in October of 1966 [24]. The campaign targeted infants at approximately age one year and all remaining susceptible children before they entered school [25, 26]. We thus assume that the individuals targeted for vaccination were children aged one year (infants), four years (pre-kindergarten), five years (kindergarten) and six years (first year of elementary school) [27]. For example, the targeted population in 1963 (column 4 of Table S1) is the total number of children born in 1962 (one year old), 1959 (four years old), 1958 (five years old), and 1957 (six years old):

$$4,167,362 + 4,295,000 + 4,255,000 + 4,308,000 = 17,025,362 \quad (\text{S19})$$

Since data on the distribution of measles vaccine by city were not available, we used the vaccination coverage calculated for the whole US (column 5 of Table S1) to estimate the proportion vaccinated in New York City (figure 3).

For the years after 1963 we had to account for vaccination being implemented already. Hence the size of the targeted population was still the total number of one, four, five and six year olds, but now reduced by the number of children already vaccinated. For example, the targeted population in the year 1964 (column 4 of Table S1) is the total number of children born in 1963 (one year old), 1960 (four years old), 1959 (five years old), 1958 (six years old) reduced by the number of children already vaccinated in 1963,  $\approx 19\%$  of five and six year olds:

$$4,098,020 + 4,257,850 + 4,295,000 + 4,255,000 - \\ -(4,295,000 + 4,255,000) \cdot 0.187955 = 15,298,855 \quad (\text{S20})$$

We then estimate the proportion vaccinated as the ratio of the distributed vaccine doses to the size of the targeted population.

In 1971 the combined measles, mumps and rubella (MMR) vaccine was licensed [29]. Since it was a new vaccine that protected against three infectious diseases, we assume that the targeted population in 1971 was again all children aged one, four, five, and six years old even if they were previously vaccinated with measles vaccine. Table S1 summarizes our calculations of the vaccination level for the US. Note that we did not consider vaccine efficacy [30], migration of the population, or other factors that could have influenced the proportion of those vaccinated who were actually immunized. Despite the limitations of our methods,

Year	Births [28]	Vaccine doses [24]	Size of targeted population	Estimated proportion vaccinated
1957	4,308,000			
1958	4,255,000			
1959	4,295,000			
1960	4,257,850			
1961	4,268,326			
1962	4,167,362			
1963	4,098,020	3,200,000	17,025,362	0.187955
1964	4,027,490	3,800,000	15,298,855	0.248385
1965	3,760,358	6,000,000	14,724,270	0.407490
1966	3,606,274	7,900,000	12,196,284	0.647738
1967	3,520,959	6,400,000	9,657,979	0.662665
1968	3,501,564	5,300,000	8,695,492	0.609511
1969	3,600,206	4,900,000	7,999,115	0.612568
1970	3,731,386	4,500,000	7,833,993	0.574420
1971	3,555,970	8,300,000	14,618,977	0.567755
1972	3,258,411	8,200,000	10,138,244	0.808819

Table S1: Estimated measles vaccination coverage in the United States.

our estimates are in excellent agreement with the only two published annual vaccination rates we have found: 61.4% in 1969 and 57.2% in 1970, as reported by the United States Immunization Survey [31].

### 3 Invariance of the period doubling bifurcation point

From the point of view of understanding transitions in measles dynamics, the most important bifurcation is the first period doubling (from annual to biennial cycles) that occurs as  $\mathcal{R}_0$  is increased above 1 (e.g., figure 9). In this appendix, we examine the precise value of  $\mathcal{R}_0$  at which this bifurcation occurs in a sequence of  $\text{SI}^n\text{R}$  and  $\text{SE}^m\text{I}^n\text{R}$  models. For a given mean generation time ( $T_{\text{gen}}$ ) and a given shape parameter of the infectious period distribution ( $n$ ), we compare the value of  $\mathcal{R}_0$  at which this important bifurcation occurs for a range of shape parameters for the latent period distribution ( $m$ ) in  $\text{SE}^m\text{I}^n\text{R}$  models. We find this value of  $\mathcal{R}_0$  is roughly invariant to  $m$  and to the existence of a latent period at all (i.e., we also compare with the  $\text{SI}^n\text{R}$  model). All of the entries in Table S2 below corresponding to points in figure 9 or figure 10.

In all the  $\text{SE}^m\text{I}^n\text{R}$  models we consider here, the mean latent and infectious periods were set to 5 and 8 days, respectively. In order to keep the mean generation time the same in all these  $\text{SE}^m\text{I}^n\text{R}$  models and in the  $\text{SI}^n\text{R}$  model, we used equation 11 to determine by exactly how much we needed to lengthen the mean infectious period ( $1/\gamma$ ) of the  $\text{SI}^n\text{R}$  model. For example, for  $n = 2$ , we first determined the mean generation time of the  $\text{SE}^m\text{I}^2\text{R}$  model as

$$T_{\text{gen}}^{\text{SE}^m\text{I}^2\text{R}} = 8 \text{ (days)} + \frac{2+1}{2 \cdot 2} \cdot 5 \text{ (days)} = 11.75 \text{ (days)}. \quad (\text{S21})$$

Then we computed the mean infectious period for the  $\text{SI}^2\text{R}$  model (equation 11 with  $1/\sigma = 0$ ), which yields a mean generation time of 11.75 days:

$$\frac{1}{\gamma_{\text{SI}^2\text{R}}} = 11.75 \text{ (days)} \cdot \frac{4}{3} \approx 15.67 \text{ (days)}. \quad (\text{S22})$$

Next we conducted a bifurcation analysis of the  $\text{SI}^n\text{R}$  model with the new values of the infectious period (figure 10).

For each  $n$ , the values of  $\mathcal{R}_0$  in Table S2 differ at most by 0.26 ( $\text{SE}^{64}\text{I}^1\text{R}$  vs  $\text{SI}^1\text{R}$ ), which strongly supports the conclusion that the  $\text{SI}^n\text{R}$  model is a good approximation of the  $\text{SE}^m\text{I}^n\text{R}$  model if the mean generation time is chosen appropriately.

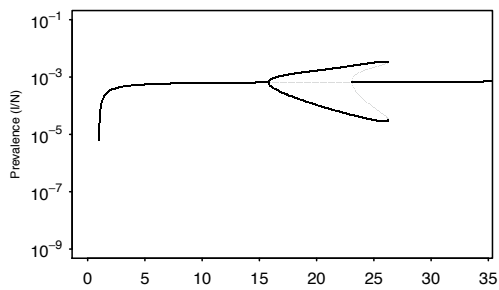
$n$	$\mathcal{R}_0$ at first period doubling bifurcation					$T_{\text{gen}}$
	$\text{SE}^1\text{I}^n\text{R}$	$\text{SE}^8\text{I}^n\text{R}$	$\text{SE}^{20}\text{I}^n\text{R}$	$\text{SE}^{64}\text{I}^n\text{R}$	$\text{SI}^n\text{R}$	
1	15.6432	15.5490	15.5418	15.5386	15.7980	13.00
2	13.9431	13.8693	13.8639	13.8613	13.9870	11.75
3	13.3869	13.3186	13.3135	13.3112	13.3999	11.33
5	12.9453	12.8811	12.8764	12.8742	12.9349	11.00
10	12.6159	12.5548	12.5503	12.5482	12.5897	10.75
20	12.4520	12.3921	12.3877	12.3857	12.4180	10.63
50	12.3537	12.2947	12.2904	12.2889	12.3151	10.55
100	12.3211	12.2622	12.2601	12.2592	12.2809	10.53

Table S2: The value of  $\mathcal{R}_0$  at the first period doubling point on the main branch of the bifurcation diagram of  $\text{SE}^m\text{I}^n\text{R}$  and  $\text{SI}^n\text{R}$  models. In each row of the table, the shape parameter  $n$  and the mean generation time are fixed. The mean generation time ( $T_{\text{gen}}$ ) can be expressed in terms of the mean latent and infectious periods using equation 11. The same formula (equation 11) with  $1/\sigma = 0$  gives the mean generation time in the  $\text{SI}^n\text{R}$  model and allows us to choose a mean infectious period ( $T_{\text{inf}} = 1/\gamma$ ) that will yield a given generation time.

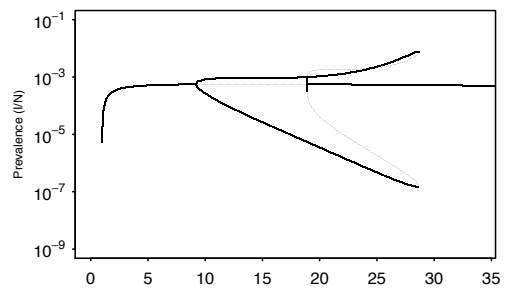
## 4 Main branch of the $SI^nR$ model with fixed mean infectious period, $\frac{1}{\gamma} = 13$ days

The sequence of graphs below show the main branch of the bifurcation tree of the  $SI^nR$  model (prevalence  $I$  vs  $\mathcal{R}_0$ ) for a sequence of shape parameters ( $n = 1, 2, \dots, 20, 30, 50, 100, \infty$ ).

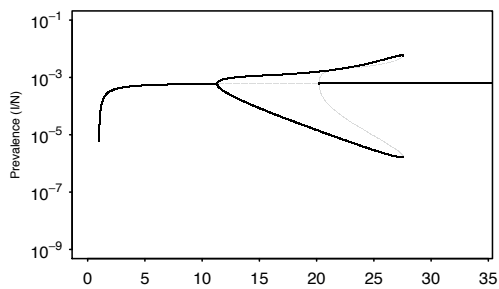
Main branch of the SI<sup>1</sup>R model



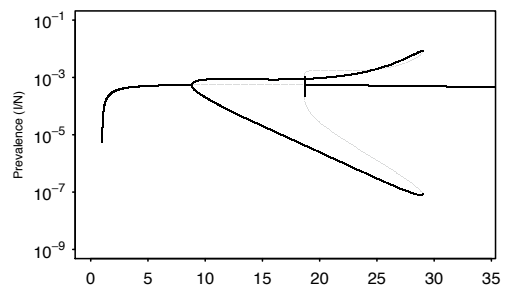
Main branch of the SI<sup>4</sup>R model



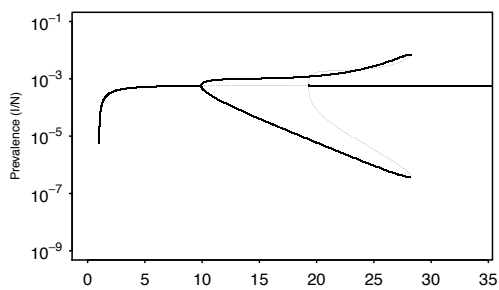
Main branch of the SI<sup>2</sup>R model



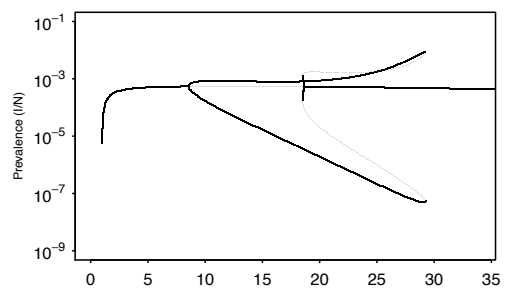
Main branch of the SI<sup>5</sup>R model



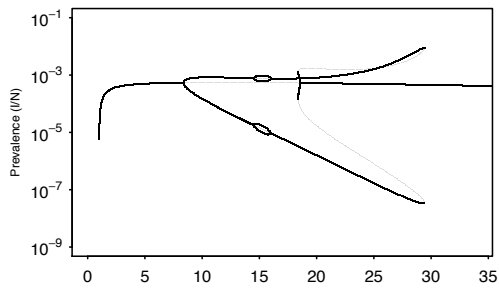
Main branch of the SI<sup>3</sup>R model



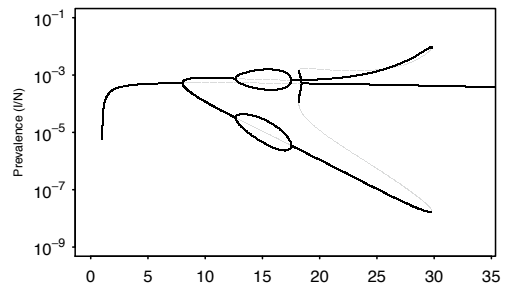
Main branch of the SI<sup>6</sup>R model



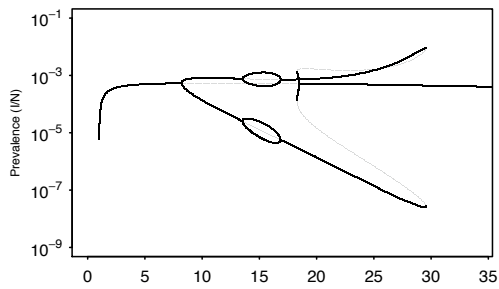
Main branch of the SI<sup>7</sup>R model



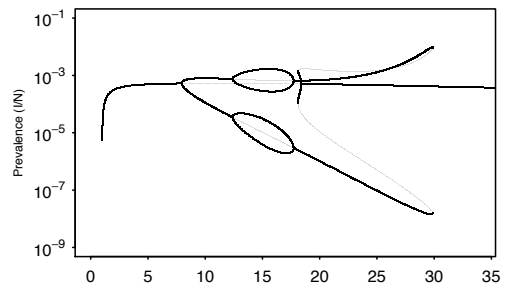
Main branch of the SI<sup>10</sup>R model



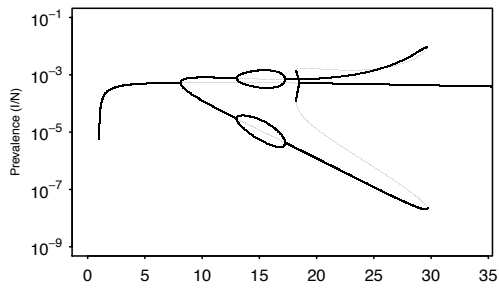
Main branch of the SI<sup>8</sup>R model



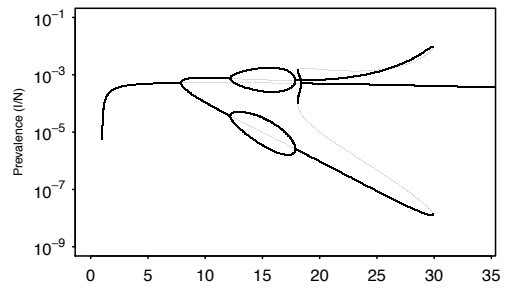
Main branch of the SI<sup>11</sup>R model



Main branch of the SI<sup>9</sup>R model

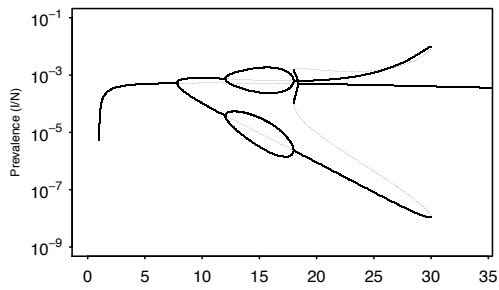


Main branch of the SI<sup>12</sup>R model

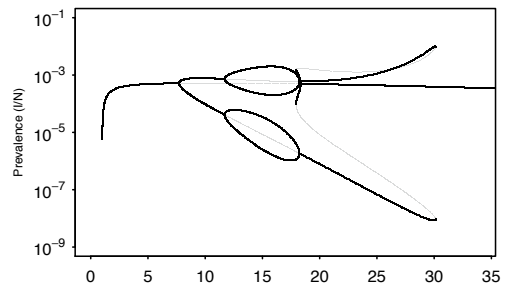




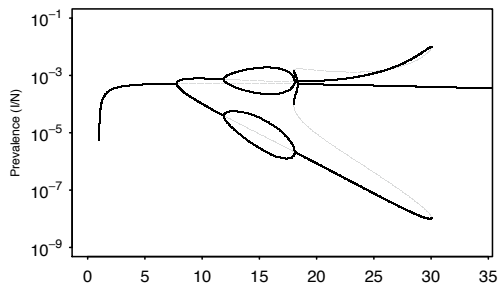
Main branch of the SI<sup>13</sup>R model



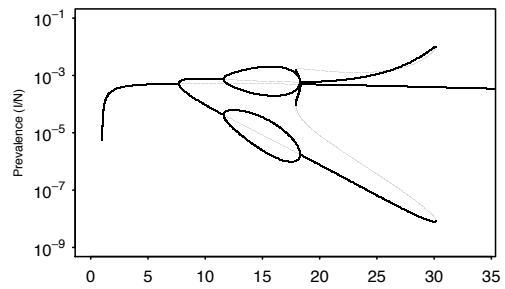
Main branch of the SI<sup>16</sup>R model



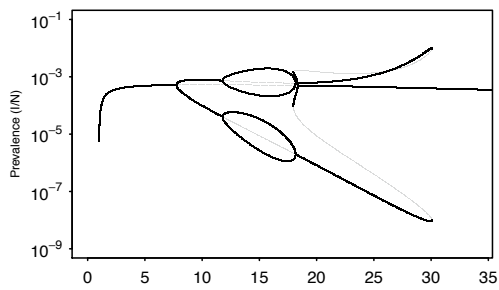
Main branch of the SI<sup>14</sup>R model



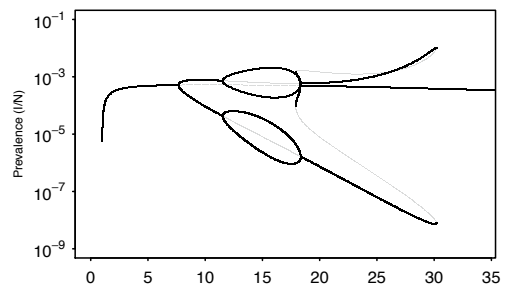
Main branch of the SI<sup>17</sup>R model



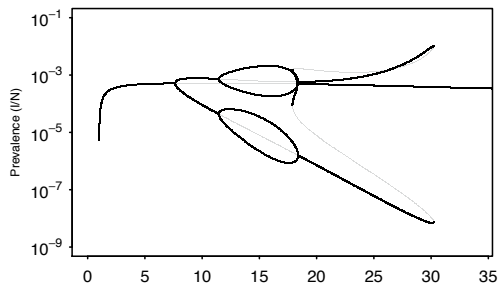
Main branch of the SI<sup>15</sup>R model



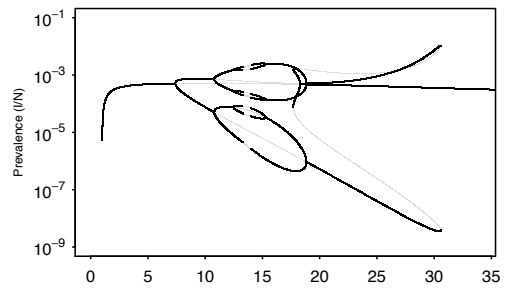
Main branch of the SI<sup>18</sup>R model



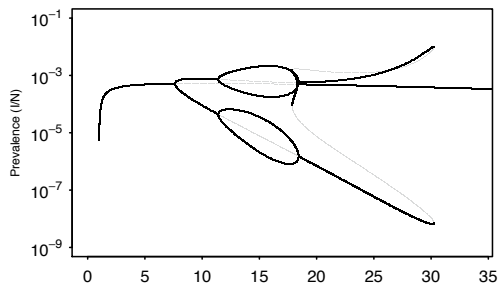
Main branch of the SI<sup>19</sup>R model



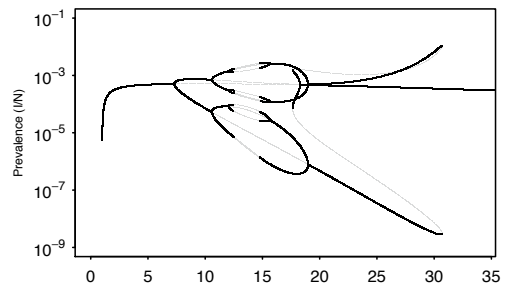
Main branch of the SI<sup>50</sup>R model



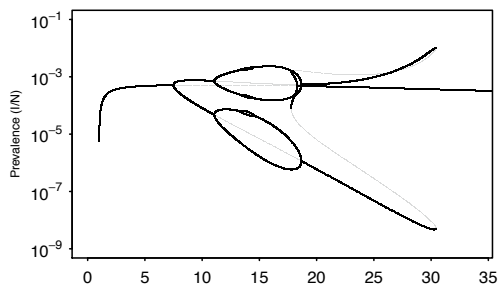
Main branch of the SI<sup>20</sup>R model



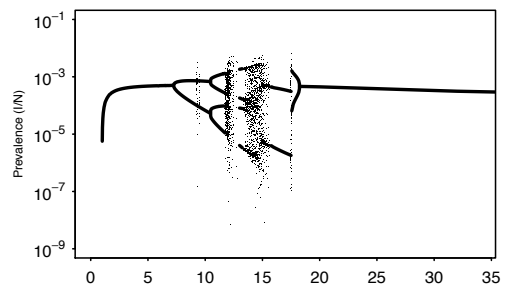
Main branch of the SI<sup>100</sup>R model



Main branch of the SI<sup>30</sup>R model



Main branch of the SI<sup>∞</sup>R model



## References

- 1 Anderson RM, May RM. Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford; 1991.
- 2 Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics. Proceedings of the Royal Society of London, Series A. 1927;115:700–721.
- 3 He D, Earn DJD. Epidemiological effects of seasonal oscillations in birth rates. Theoretical Population Biology. 2007;72(2):274–291.
- 4 Earn DJD. Mathematical epidemiology of infectious diseases. In: Lewis MA, Chaplain MAJ, Keener JP, Maini PK, editors. Mathematical Biology. vol. 14 of IAS/ Park City Mathematics Series. American Mathematical Society; 2009. p. 151–186.
- 5 Earn DJD, Rohani P, Bolker BM, Grenfell BT. A simple model for complex dynamical transitions in epidemics. Science. 2000;287(5453):667–670.
- 6 Bauch CT, Earn DJD. Interepidemic intervals in forced and unforced SEIR models. Fields Institute Communications. 2003;36:33–44.
- 7 Bauch CT, Earn DJD. Transients and attractors in epidemics. Proceeding of the Royal Society of London Series B-Biological Sciences. 2003;207(1524):1573–1578.
- 8 Ma J, Earn DJD. Generality of the final size formula for an epidemic of a newly invading infectious disease. Bulletin of Mathematical Biology. 2006;68(3):679–702.
- 9 Hethcote HW, Tudor DW. Integral equation models for endemic infectious diseases. Journal of Mathematical Biology. 1980;9(1):37–47.
- 10 Feng Z, Thieme HR. Endemic models with arbitrarily distributed periods of infection I: fundamental properties of the model. SIAM Journal of Applied Mathematics. 2000;61(3):803–833.
- 11 Lloyd AL. Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. Proceedings of the Royal Society of London, Series B. 2001;268(1470):985–993.

- 12 Bailey N. Some stochastic models for small epidemics in large populations. *Applied Statistics*. 1964;13(1):9–19.
- 13 Anderson D, Watson R. On the spread of a disease with gamma distributed latent and infectious periods. *Biometrika*. 1980;67(1):191–198.
- 14 Therrien C, Tummala M. *Probability for Electrical and Computer Engineers*. 2nd ed. Taylor and Francis; 2011.
- 15 Feng ZL, Xu DS, Zhao HY. Epidemiological models with non-exponentially distributed disease stages and applications to disease control. *Bulletin of Mathematical Biology*. 2007;69(5):1511–1536.
- 16 Keeling MJ, Grenfell BT. Understanding the persistence of measles: reconciling theory, simulation and observation. *Proceedings of the Royal Society of London Series B-Biological Sciences*. 2002;269(1489):335–343.
- 17 Nguyen HTH, Rohani P. Noise, nonlinearity and seasonality: the epidemics of whooping cough revisited. *Journal of the Royal Society Interface*. 2008;5(21):403–413.
- 18 Grossman Z. Oscillatory phenomenon in a model of infectious diseases. *Theoretical Population Biology*. 1980;18(2):204–243.
- 19 Olsen LF, Schaffer WM. Chaos versus noisy periodicity: alternative hypotheses for childhood epidemics. *Science*. 1990;249(4968):499–504.
- 20 Aron JL, Schwartz IB. Seasonality and period-doubling bifurcations in an epidemic model. *Journal of Theoretical Biology*. 1984;110(4):665–679.
- 21 Schenzle D. An age-structured model of pre- and post-vaccination measles transmission. *IMA Journal of Mathematics Applied in Medicine and Biology*. 1984;1(2):169–191.
- 22 Finkenstadt BF, Grenfell BT. Time series modelling of childhood diseases: a dynamical systems approach. *Journal of the Royal Statistical Society Series C - Applied Statistics*. 2000;49(2):187–205.
- 23 Measles, Mumps, Rubella; 2010. National Network for Immunization Information. Available from: <http://www.immunizationinfo.org/vaccines/measles>.

- 24 Witte JJ, Axnick NW. The benefits from 10 years of measles immunization in the United States. *Public Health Reports*. 1975;90(3):205–7. 807933.
- 25 Dull HB, Witte JJ. Progress of measles eradication in the United States. *Public Health Reports*. 1968;83(3):245–248.
- 26 Conrad JL, Wallace R, Witte JJ. The epidemiologic rationale for the failure to eradicate measles in the United States. *American Journal of Public Health*. 1971;61(11):2304–2310.
- 27 Structure of US Education; 2011. US Department of State. Available from: <http://infousa.reingoldweb.com/education/overview/edlite-structure-us.html>.
- 28 Vital Statistics of the United States; 2011. Center for Disease Control and Prevention. National Center for Health Statistics. Available from: <http://www.cdc.gov/nchs/products/vsus.htm>.
- 29 Banatvala JE, Brown DWG. Rubella. *Lancet*. 2004;363(9415):1127–1137.
- 30 Linnemann CCJ. Measles vaccine: immunity, reinfection and revaccination. *American Journal of Epidemiology*. 1973;97(6):365–371.
- 31 Landrigan PJ, Conrad JL. Current Status of Measles in the United States. *The Journal of Infectious Diseases*. 1971;124(6):620–622.

Bifurcation analysis of the  
seasonally forced SIR model  
using XPPAUT

*Supplementary Material for*  
“Effects of the infectious period distribution  
on predicted transitions in childhood disease  
dynamics”

Olga Krylova and David J.D. Earn

*Department of Mathematics and Statistics, McMaster University,  
Hamilton, Ontario, Canada L8S 4K1*

April 14, 2013


**Abstract**

This guide provides a recipe for constructing bifurcation diagrams of seasonally forced epidemic models using XPPAUT, with which the reader is assumed to have basic familiarity. A preliminary version of this document appeared in Olga Krylova’s McMaster University PhD thesis [1]. It extends part of a set of unpublished notes written by Victoria Maystruk (2006) as a supplement to her McMaster University Undergraduate Arts & Science thesis [2].

We are continuing to improve this guide. Please e-mail David Earn ([earn@math.mcmaster.ca](mailto:earn@math.mcmaster.ca)) if you notice any errors or have suggestions for improvements. In your message, please indicate the date on the version of the guide that you have been reading.

# 1 Preliminaries

This guide was prepared on a Macintosh computer running:

- Mac OS X version 10.6.8
- XPPAUT version 6.11
-  version 2.15.2 (2012-10-26) and **knitr** package version 1.1
- gcc version 4.2.1

The “command line” refers to the unix command line, accessible from the Terminal application under Mac OS X.

# 2 The model

Consider a seasonally forced SIR model:

$$\frac{dS}{dt} = \mu - \beta(t)S(t)I(t) - \mu S(t), \quad (1a)$$

$$\frac{dI}{dt} = \beta(t)S(t)I(t) - \gamma I(t) - \mu I(t), \quad (1b)$$

$$\frac{dR}{dt} = \gamma I(t) - \mu R(t), \quad (1c)$$

where  $S$ ,  $I$ , and  $R$  are proportions of the population in each epidemiological state and the seasonally forced transmission rate is

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

We want to investigate how the dynamics of system (1) change with respect to changes in the basic reproduction number,  $\mathcal{R}_0$ . Thus we write equation 1 in terms of  $\mathcal{R}_0$ :

$$\mathcal{R}_0 = \frac{\langle \beta \rangle}{\gamma + \mu}, \quad (3)$$

which implies that

$$\langle \beta \rangle = \mathcal{R}_0(\gamma + \mu), \quad (4a)$$

$$\beta(t) = \mathcal{R}_0(\gamma + \mu)(1 + \alpha \cos(2\pi t)). \quad (4b)$$

A bifurcation diagram is a summary of the asymptotic dynamics (attractors) of a dynamical system as a function of a bifurcation parameter. It is helpful to start analysis of a model with a brute force bifurcation diagram, which is constructed strictly by simulations. This diagram will show stable branches of periodic orbits and will identify regions of possible chaotic behavior.

## 3 STEP 1: Brute force bifurcation diagram

### 3.1 XPPAUT input file *bruteforceSIR\_R0.ode*

In order to compute solutions of any model in XPPAUT we need to create a `filename.ode` file that specifies the system of differential equations, initial conditions, parameter values and any output that we want to save. The `.ode` file can be created in any plain text editor (*e.g.*, Emacs, vim or TextEdit).

**Dynamical equations:** Equations (1) are specified as follows.

```
## DIFFERENTIAL EQUATIONS:
beta=Rzero*(gamma+mu)*(1 + a*cos(2*pi*t))
s'=mu-beta*s*i-mu*s
i'=beta*s*i-(gamma+mu)*i
```

Using this much of `.ode` file we can explore solutions of the SIR model for various initial conditions and parameter values interactively in the XPPAUT window (see [3] for detailed explanations of working with XPPAUT).

**Initial conditions and parameter values:** To set these in advance so we can run XPPAUT jobs in batch mode, we use `init` and `par` lines in the `.ode` file.

```
## INITIAL CONDITIONS:
init S=0.9, I=0.001

## PARAMETERS:
## mean lifetime 1/mu = 50 years
## mean infectious period 1/gamma = 13 days = 0.0356 years
## amplitude of seasonality alpha = 0.08
par mu=0.02, gamma=28.08, Rzero=17, a=0.08
```

Our goal is to plot the solution of the system for a given range of values of the bifurcation parameter,  $\mathcal{R}_0$ , on the horizontal axis and a particular state variable on the vertical axis,  $\log_{10} I$ . We choose  $\log_{10} I$  instead of  $I$  since the proportion infectious becomes very small, and it is convenient to use a logarithmic scale for a better visual representation of the results. In order to save the value of our bifurcation parameter ( $\mathcal{R}_0$ ) and the logarithms of the state variables, we must specify *auxilliary variables* in our `.ode` file.



**Auxiliary variables:** Additional quantities defined in terms of the state variables and/or parameters are saved by XPPAUT if specified as follows. We save  $\mathcal{R}_0$  because we want to plot a bifurcation diagram as a function of  $\mathcal{R}_0$ . XPPAUT does not save parameter values by default, so we need to create auxiliary variables for any parameters we need to track.

```
## AUXILIARY VARIABLES:  
aux R0=Rzero  
aux log10s=log10(s)  
aux log10i=log10(i)
```

**XPP setup:** When using our .ode file for interactive exploration with XPPAUT, it is convenient to have some default plotting options set. This will have no effect on the output XPPAUT generates when run in batch (“silent”) mode.

```
## PLOT OPTIONS:  
## xp=variable on x axis, yp=variable on y axis  
@ xp=R0, yp=log10i  
## limits on plot  
@ xlo=0, xhi=40, yhi=0, ylo=-25  
## background colour for plot  
@ back=white
```

**Poincaré map.** The seasonally forced SIR model exhibits periodic solutions, so if we plot the complete solution for  $I(t)$  for each particular value of  $\mathcal{R}_0$  we will get straight vertical intervals corresponding to the range of the  $I(t)$  at a particular  $\mathcal{R}_0$ . Such a diagram would be more confusing than useful. It is more convenient to use the Poincaré stroboscopic map. Choosing the strobing interval to be one year, the Poincaré map will return just one point for a period-one orbit, two points for a period-two orbit and so on. This representation of the results is much more intuitive since it allows us to show all periodic solutions of the model and their periods. In XPPAUT, the Poincaré map can be set up in the following way:

```
## POINCARÉ MAP SET UP:  
@ poimap=section,poivar=t,poiplt=1
```

The range of our bifurcation parameter ( $\mathcal{R}_0$ ) is set as follows.

```
## range set up
@ range=1, rangeover=Rzero, rangestep=3000
@ rangelow=0, rangehigh=30, rangereset=no
```

**Numerical integration settings.** Since our aim to identify the *attractors* of the dynamical system, we must ignore transient dynamics before convergence onto an attractor. To this end, we choose the total integration time to be 650 years and ignore the first 600 years, which is sufficient for solutions to converge to an attractor.

```
## INTEGRATION OPTIONS:
## total time of integration
@ total=650,
## transient time
@ trans=600
## time step for integration
@ dt=0.001
```

**Storage and data saving settings.** The default maximum number of time points to store is 5000, which is not sufficient for our purposes.

```
## STORAGE and DATA SAVING OPTIONS
## max number of time points to store (default 5000)
@ maxstor=2000000
## filename for output to be saved
@ output=bruteforceSIR_R0.dat
done
```

**Running a job with XPPAUT.** Our `.ode` file can be executed from the command line as a batch job using “silent mode” as follows:

```
xppaut bruteforceSIR_R0.ode -silent
```

Alternatively, this can be done from within an  $\mathbb{R}$  script via

```
system("xppaut bruteforceSIR_R0.ode -silent")
```

Whether executed from the command line or from within  $\mathbb{R}$ , this will produce a `bruteforceSIR_R0.dat` file with all the necessary data:

column 1	$time$
column 2	$S$
column 3	$I$
column 4	$\mathcal{R}_0$
column 5	$\log_{10} S$
column 6	$\log_{10} I$

Generating this file took 80 CPU minutes on a Macintosh computer with a 2.53 GHz Intel Core 2 Duo processor.

## 3.2 Plotting the brute force bifurcation diagram

We begin by reading the brute force data into a data frame (`bfd`).

```
bfd <- read.table("bruteforceSIR_R0.dat",  
  col.names=c("time", "S", "I", "R0", "log10S", "log10I"))
```

To avoid spurious detection of long period cycles, we specify that maximum period of interest to us.

```
maxper <- 7
```

**Function to calculate the period of the attractor for each  $\mathcal{R}_0$ .** We would like to colour branches according to the period of the attractor. The following function creates a data frame containing only the last point on each trajectory for each  $\mathcal{R}_0$  value, and includes a `period` column. The arguments are

- df** the original data frame,
- dop** the digits of precision to be used for cycle detection,
- R0lim** an “ $\mathcal{R}_0$  limit”, used to specify that the period is actually 1 for all  $\mathcal{R}_0$  less than this limit, even though a longer period is detected due to slow convergence for low  $\mathcal{R}_0$ ,

**max.period** periods above this will be set to NA since they presumably result from slow convergence.

```
last.point.with.period <- function(df, dop=5,
                                R0lim=4, max.period=maxper) {
  ## data frame with only the last pt on each soln:
  df.last <- subset(df, time==max(time))
  nR0 <- nrow(df.last) # number of R0 values
  df.last$period <- rep(0,nR0) # add period column
  for (i in 1:nR0) {
    ## data frame with all pts on soln with given R0:
    R0i <- df.last[i,"R0"]
    df.R0i <- subset(df,R0==R0i)
    ## compute period of this solution:
    if (R0i < R0lim) {
      period <- 1
    } else {
      period <- length(unique(round(df.R0i[, "log10I"], dop)))
      if (period > max.period) period <- NA
    }
    df.last$period[i] <- period
  }
  return(df.last)
}
```


**Function to plot period of attractor as a function of  $\mathcal{R}_0$ .** Given a data frame that contains a period column, we can easily plot the periods as a function of  $\mathcal{R}_0$ .

```
plot.period.of.R0 <- function(df, R0lim=1, ...) {
  with(df, {
    period[which(R0 < R0lim)] <- 1
    plot(R0, period, pch=".", col=period, cex=period,
         xlab=expression(R[0]),
         ylab="Period [yr]",
         cex.axis=1.3, cex.lab=1.5, las=1,
```

```

    ... )
  })
}

```

**Function to plot brute force bifurcation diagram.** We use the following  function to plot figure 2, which clearly shows the existence of periodic orbits of periods one to seven. Having discovered that the longest cycle is seven years, we filter out only the last seven time steps of data for each  $\mathcal{R}_0$  value. This greatly reduces the number of points plotted (and hence the pdf file size) without making any difference to the appearance of the plot. The only required argument of the function is the data frame containing the XPPAUT output.

```

bruteforceplot <- function(df, tsave=maxper, R0lim=1,
                           xlim=c(0,33), ...) {
  df <- subset(df, time>max(time)-tsave)
  dflp <- last.point.with.period(df)
  df$period <- rep(0,nrow(df)) # add period col to df
  for (i in 1:nrow(df)) {
    R0i <- df[i,"R0"]
    df$period[i] <- if (R0i > R0lim)
      dflp[dflp$R0==R0i,"period"] else 1
  }
  with(df,{
    par(mar=c(5,5,0,2)) # alter margins
    xlab <- expression(paste(
      "Basic Reproduction Number, ",R[0]))
    ylab <- "Prevalence, I/N"
    plot(R0, log10I, pch=".", yaxt="n", xaxs="i",
          xlim=xlim, las=1, xlab=xlab, ylab=ylab,
          cex.axis=1.3, cex.lab=1.5,
          col=period, cex=period, ...)
    ## add y-axis annotation
    y.ticks <- 0:-8
    y.label <- paste("10", y.ticks, sep="^")
    axis(2, at=y.ticks, las=2, label=parse(text=y.label))
  })
}

```

```
}
```

**Brute force bifurcation plots.** Figure 1 displays the simple plot showing the periods of the detected attractors. Note that near the period doubling bifurcation, convergence to the attractor is slower (hence the two spurious detections of longer period cycles near that bifurcation point). Figure 2 shows the more traditional bifurcation diagram.

```
bfd.last <- last.point.with.period(bfd)
plot.period.of.R0(bfd.last)
```

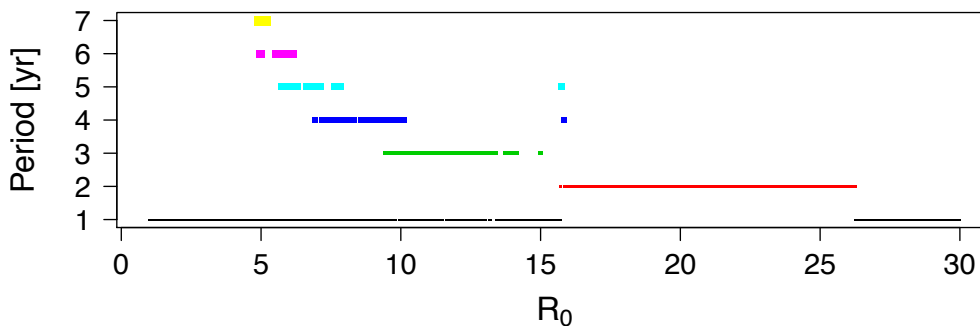



Figure 1: Periods of detected attractors for the sinusoidally forced SIR model (1) with mean infectious period  $1/\gamma = 13$  days, mean lifetime  $1/\mu = 50$  years and amplitude of seasonality  $\alpha = 0.08$ .

## 4 Save a set of initial conditions for use in AUTO

The following  code reads the data from the previously generated brute force bifurcation data file and saves a set of initial conditions that will be used to find each periodic orbit in AUTO.

```
bruteforceplot(bfd, ylim=c(-8,-2))
```

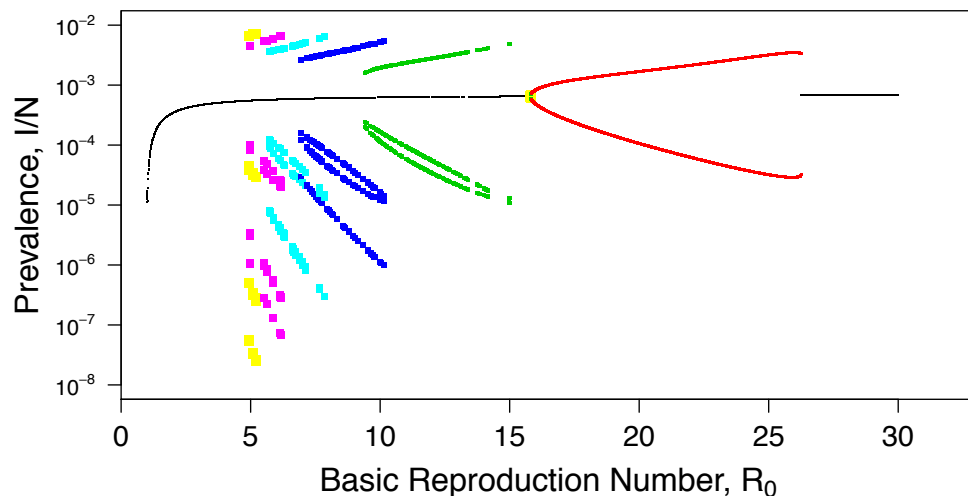


Figure 2: Brute force bifurcation diagram for the sinusoidally forced SIR model (1) with mean infectious period  $1/\gamma = 13$  days, mean lifetime  $1/\mu = 50$  years and amplitude of seasonality  $\alpha = 0.08$ .

**For each period that occurs, save a single point to be used as an initial condition in AUTO.** For each period, we save the final condition for the “middle”  $\mathcal{R}_0$  value associated with this period, to avoid bifurcation points and slowly converging solutions at the “edges”. Note that if there are multiple branches with the same period, only one branch will be saved.

```
## extract all periods that are not NA:
all.periods <- with(bfd.last, period[!is.na(period)])
(unique.periods <- unique(all.periods))

## [1] 1 7 6 5 4 3 2

nper <- length(unique.periods)
## Create data frame for list of initial conditions:
ic.set <- bfd.last[1,]
## For each period that occurs, save a final condition:
for (iper in 1:nper) {
  ## select the rows in the data frame with this period:
```

```

df.iper <- subset(bfd.last, period==unique.periods[iper])
## choose the middle row to avoid bifurcation points:
ic.set[iper,] <- df.iper[round(mean(1:nrow(df.iper))),]
}
## replace original row names with ic number:
row.names(ic.set) <- 1:nrow(ic.set)
## Save these points to be used as initial conditions:
write.csv(ic.set, "icset.csv",
          row.names=FALSE, quote=FALSE)
save(ic.set, file="icset.rda")

```

**Table of initial conditions for use in AUTO.** The structure of `setIC.csv` is

time	$S$	$I$	$\mathcal{R}_0$	$\log_{10} S$	$\log_{10} I$	period
639	0.038555704	0.00067852496	27.059999	-1.4139113	-3.1684341	1
647	0.056224938	0.0016748	19.950001	-1.250071	-2.776037	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮

The actual list of initial conditions that we have generated is:

```

print(subset(ic.set, select = c("R0", "period", "S", "I")))
##      R0 period      S      I
## 1  9.58      1 0.10549 6.190e-04
## 2  5.10      7 0.23534 3.180e-05
## 3  5.64      6 0.17693 2.196e-07
## 4  6.62      5 0.17512 3.322e-05
## 5  8.18      4 0.13811 3.765e-05
## 6 11.83      3 0.07438 6.034e-05
## 7 21.05      2 0.05360 1.908e-03

```

## 5 STEP 2: AUTO bifurcation diagram

XPPAUT provides an interface to AUTO, which allows us to track bifurcation curves for steady-state and periodic systems [4]. In particular, we can use AUTO to follow stable and unstable branches of periodic orbits. Unfortunately, AUTO's



default settings only allow us to plot the maximum, minimum, mean or period of the cycle. It is easier to visualize the bifurcation structure when plotting the diagram as a Poincaré map, similar to the brute force diagram shown in figure 2. This way, annual cycles will be displayed with a single solid line, biennial cycles with a double line, triennial cycles with a triple line, *etc.* [5, 6]. Consequently, we choose to use AUTO to study the discrete time system given by the Poincaré map (one-year stroboscopic map), associated with our original continuous-time system. We can analyze the discrete-time system, keeping in mind that a fixed point of a Poincaré map corresponds to a periodic orbit of the original system. The problem of stability of the periodic orbits of the original system is reduced to the problem of stability of fixed points of maps. Stability is determined by the eigenvalues of the map linearized about the fixed points.

## 5.1 Constructing the Poincaré map using a C-library

We construct our stroboscopic map by linking XPPAUT to a C function that numerically integrates the original continuous-time system for a specified period of time (one year in our case). The numerical integration can be done by any method we choose (*e.g.*, Euler, Runge-Kutta or something more sophisticated). Our C function uses initial conditions specified in our `.ode` file or in the XPP window and returns the integrated values after one year as output. The file `SIRmap.c` below is compiled with a C compiler and stored in a shared object library, to which we link from our `AutoSIR.ode` file (§5.2) where all initial values and parameters are specified.

```

/**
 Poincare map of the seasonally forced SIR model
 to be called from XPPAUT for bifurcation analysis.

 Under MacOSX, compile this function via:
 gcc -dynamiclib -m32 -o SIRmap.so SIRmap.c
 ***/

#include <math.h>

/** Compile-time definitions ***/
#define Time_step 0.0005 /* in units of years */
#define TWO_PI 6.2831853071795864769252867665590

```

```

#define Real double
#define NDIM 2 /* dimension of the dynamical system */

/* **** */
/* ***/ FUNCTIONS CALLED BY THE MAIN ROUTINE ***/
/* **** */

/* ***/ Euler integrator ***/
void Euler(Real *x, Real *dx, Real dt, int ndim ) {
    int i;
    for (i=0; i<ndim; i++) {
        x[i] = x[i] + dx[i]*dt;
    }
}

/* ***/ Seasonally forced transmission rate ***/
Real Seasonal_beta(Real beta0, Real alpha, Real time) {
    Real c2pt; /* cos(2*pi*t) */

    c2pt = cos(TWO_PI*time);
    return(beta0*(1+alpha*c2pt));
}

/* **** */
/* ***/ THE MAIN ROUTINE ***/
/* **** */

/* ***/ The function SIRmap is what XPPAUT calls ***/
SIRmap(Real *in, Real *out, int nin,
        int nout, Real *var, Real *con)
/*
    in = initial and parameter values we get
    from the ode file (s, i, R0, alpha, gamma, mu)
    out = what we are returning (sp, ip):
    calculated values of S and I after one year
    nin = dimension of in[]
    nout = dimension of out[]
*/

```

{

```
/* define starting values in log base 10 */
Real s=in[0], i=in[1];
Real x[NDIM], dx[NDIM]; /* for Euler integrator */
/* converting back to the original values,
   not in log */
s=pow (10,s);
i=pow (10,i);
/* define parameter values */
Real R0=in[2], alpha=in[3], gamma=in[4], mu=in[5];
Real ds, di;
Real beta0, nonlin_term;
Real time; /* in units of years */
long istep, nsteps;

/* number of steps in a year */
nsteps = (int)( 1/Time_step + 0.5 );

/* integrating for one year */
for (istep=0; istep < nsteps; istep++) {
    time=(Real)(istep)*Time_step;

    /* compute the vector field */
    beta0 = R0*(gamma+mu); /* mean transmission rate */
    nonlin_term = Seasonal_beta(beta0, alpha, time) * s * i;

    ds = mu - nonlin_term - mu*s;
    di = nonlin_term - (mu + gamma)*i;

    /* integrate using euler's method */
    x[0] = s; x[1] = i; dx[0]=ds; dx[1]=di;
    Euler(x, dx, Time_step, NDIM);
    s = x[0]; i = x[1];
}
s = log10(s);
i = log10(i);
out[0] = s; /* Output in log_10 */
out[1] = i;
```

```
}
```

## 5.2 Creating file *AutoSIR.ode*

This file links our C-library function `SIRmap` to XPPAUT so that we can conduct bifurcation and continuation analysis of the seasonally forced SIR model interactively from the AUTO window in XPPAUT. The equations are defined in the `.ode` file using the approach discussed in §9.8 of Ermentrout's book [3, p.250].

```
## SPECIFY EQUATIONS (cf. Ermentrout 2002, p.250)
s' = sp
i' = ip
sp = 0
ip = 0
```

**Setting up the interaction with the C-library.** In our `.ode` file, the command that links to our C function `SIRmap` in `SIRmap.c` has the form:

```
export {in} {out}
```

where `{in}` indicates initial values and parameters to be passed to the external function and `{out}` defines returning values.

```
## LINK TO THE C-LIBRARY
## We pass the values {s0, i0, R0, alpha, gamma, mu}
## and ask the C-library function to return {sp, ip}.
## The order of export must agree with the order of
## the in[] and out[] arrays in the C function.
export {s, i, R0, alpha, gamma, mu} {sp, ip}
## Specify the name of the shared object library, and
## the name of the function within it that we need.
@ dll_lib=SIRmap.so dll_fun=SIRmap
```

**Define set of initial conditions for each periodic orbit.** XPPAUT allows us to pre-define a set of initial conditions and parameter values in an `.ode` file. We do this here for initial conditions saved from the brute force bifurcation diagram. The general format of the command in the `.ode` file is

```
set name {parameter values, initial data, options}
```

These sets can be invoked while running XPPAUT interactively with the

```
File→Get par set
```

command. For example, this set yields a period 1 (annual) cycle:

```
set p1 {init s=-1.41, init i=-3.17, R0=27.06, nout=1}
```

Note here that `nout` refers to the number of iterations of the map between outputs [3, p.25], *i.e.*, this is the period of the periodic orbit in question. In discrete time, XPPAUT is designed to analyze equilibria only, so by setting `nout` to an integer larger than 1, we “trick” XPPAUT into analyzing a periodic orbit of period greater than 1.

```
## SET INITIAL CONDITIONS for one point on each
## periodic orbit of interest.
## For historical reasons, s and i below actually
## refer to log10(s) and log10(i).
set p1 {init s=-1.41, init i=-3.17, R0=27.06, nout=1}
set p2 {init s=-1.25, init i=-2.78, R0=19.95, nout=2}
set p3 {init s=-1.16, init i=-4.39, R0=12.58, nout=3}
set p4 {init s=-0.91, init i=-4.82, R0= 9.53, nout=4}
set p5 {init s=-0.74, init i=-5.14, R0= 5.75, nout=5}
set p6 {init s=-0.71, init i=-2.27, R0= 5.53, nout=6}
set p7 {init s=-0.70, init i=-2.14, R0= 5.19, nout=7}
```

**Set default parameter values.** The order in which parameters are listed in the `par` line below determines the “main” versus “secondary” parameters for AUTO. The “main” parameter will be used for one-dimensional bifurcation diagrams. The secondary parameter listed will be used as well for two-dimensional bifurcation diagrams. For convenience, we save the value of  $\mathcal{R}_0$  in an auxiliary variable (which will be written by XPPAUT to the output file).

```

## PARAMETER VALUES
## R0 = basic reproduction number
## alpha = amplitude of seasonal forcing
## gamma = recovery rate; 1/gamma = mean infectious period
## mu = birth and death rates; 1/mu = mean lifetime
par R0=30, alpha=0.08, gamma=28.076923, mu=0.02
aux Rzero=R0

```

**XPP setup.** We need to specify that we will be iterating a discrete map as opposed to an ODE. We also set XPP graphical parameters in case we use the .ode file for exploration of solutions using XPP.

```

## XPP SETUP
## this is a discrete map not an ODE
@ meth=discrete
## total=20 means 20 iterations of the map in total
@ total=20, yp=i
## line type = dots
@ lt=0
## plotting options
@ xlo=-1, xhi=21, ylo=-9, yhi=-1

```

**AUTO setup.** It is convenient to set all necessary AUTO plotting options in the ODE file.

```

## AUTO PLOT SETTINGS
## set range for R0, our control parameter:
@ parmin=1.1, parmax=32
## set range of vertical axis variable and which
## variable it is (note that i refers to log10(i)).
@ autoymin=-9, autoymax=-1, autovar=i
## set horizontal axis plot range:
@ autoxmin=0, autoxmax=32

```

We also set parameters that control AUTO's continuation algorithm.

```

## set step size for continuation of the control parameter:
## (ds=standard step size, others are max and min step size)
## (the sign of ds controls the direction of continuation)
@ ds=0.003, dsmin=0.0000003, dsmax=0.1
## set a few other technical aspects of the continuation:
## Nmax = maximum number of steps to take along a branch
##       before stopping
## Npr  = number of steps before labelling a point
##       (which can help with continuing from points
##       without having to start from scratch)
## eps... = various tolerances
@ Nmax=20000, Npr=2000, epsl=1e-6, epsu=1e-6, sepss=1e-4
done

```

### 5.3 Following periodic orbits

We have everything ready to compute the bifurcation diagram in AUTO.

- Run XPPAUT
  - `xppaut AutoSIR.ode`
- Find period one orbit
  - Choose the first set of initial values called `p1`
    - `(F)ile→(G)et par set`
  - Integrate the system to reach an equilibrium
    - `(I)nitialconds→(G)o`

We should get 20 points in a straight horizontal line, indicating that the system is at equilibrium. To assure that the system indeed converged to an equilibrium, check data values by clicking the `Data` button on the top panel of the XPP window. Data values of  $S$  and  $I$  are still different after the third decimal place, which shows that we have not converged to the fixed point yet. Hence we need to integrate the system a few more times by pressing `(I)nitialconds→(L)ast` until we get repeating values for  $S$  and  $I$ . After repeating `I L` two more times we find an equilibrium.

- Load AUTO
  - `(F)ile→(A)uto`
- Continue the branch in both directions

- Hit Run to follow the period one branch in a positive direction ( $\mathcal{R}_0$  increases).
- Change the sign of parameter  $D_s$  in the Numerics menu of AUTO to continue computations in the opposite direction ( $\mathcal{R}_0$  decreases). Select the initial point with Grab and press Run again.
  - Numerics  $\rightarrow$   $D_s$  : -0.003
  - Grab  $\rightarrow$  <Return>  $\rightarrow$  Run

We have computed the period one orbit.

- Save data
  - File  $\rightarrow$  All info  $\rightarrow$  name file branch1allinfo.dat
- Clear memory for the new branch
  - File  $\rightarrow$  Reset diagram  $\rightarrow$  hit OK and erase data
- Compute period two branch
  - Integrate the system to reach an equilibrium
    - F  $\rightarrow$  G  $\rightarrow$  choose set p2  $\rightarrow$  I  $\rightarrow$  G  $\rightarrow$  I  $\rightarrow$  L  $\rightarrow$  I  $\rightarrow$  L ...
    - Integration results in two horizontal dotted lines indicating that we hit a period two cycle. Check the Data window to verify that the values of  $S$  and  $I$  oscillate between two constant values.
  - Define the period of the orbit. To continue branches of periodic points of the Poincaré map we have to indicate how many times the map should be iterated before returning the value of the map. This can be done using parameter  $n(O)_{\text{output}}$  from the  $n(U)_{\text{merics}}$  menu. For example, setting  $n(O)_{\text{output}}$  to two allows us to see all of the branches up to period two. Note that the previous value of  $n(O)_{\text{output}}$  was one, which allowed us to plot only period one orbits.
    - $n(U)_{\text{merics}} \rightarrow n(O)_{\text{output}} \rightarrow$  type 2  $\rightarrow$  <Return>
  - Run AUTO
    - F  $\rightarrow$  A  $\rightarrow$  Run
    - We want to have separate data files for each periodic orbit, so that we can plot them later in different colours. Therefore, when computing period 2 orbit, we have to hit the ABORT button very quickly to avoid plotting period one points.
  - Save data
    - File  $\rightarrow$  All info  $\rightarrow$  name file branch2allinfo.dat
- Compute period three branch
  - Integrate the system to reach an equilibrium
    - F  $\rightarrow$  G  $\rightarrow$  choose set p3  $\rightarrow$  I  $\rightarrow$  G  $\rightarrow$  I  $\rightarrow$  L  $\rightarrow$  I  $\rightarrow$  L ...
    - Three horizontal dotted lines are plotted in the XPP window indicating



that we hit a period three cycle. In the `Data` window, values of  $S$  and  $I$  oscillate between three constant values.

- Define the period of the orbit

`n(U)mericals → n(O)utput → type 3 → <Return>`

- Run AUTO

`F → A → Run`

You will see that AUTO computed only one of the three branches of the period three orbit. Since the branches of the period three orbit are disconnected (we see that from the brute force diagram), AUTO can detect only a branch based on the initial conditions. Hence to compute the other two branches we must start running AUTO from points on these branches. This can be achieved by using `Data` recorded in the main XPP window from our integrations. Back in the main XPP window, click the `Data` button. That window contains three distinct values of  $S$  and  $I$ . To set one of them as the initial point, make that line the first line in the XPP `Data` window and press `Get`. While this operation will assign appropriate values to  $S$  and  $I$ , the value of the bifurcation parameter  $\mathcal{R}_0$  will not be changed accordingly and that must be done manually. Click on the `Param` button in the XPP top menu to change the value of  $\mathcal{R}_0$  to the one saved under the column `Rzero` in the `Data` window. Do not forget to click `ok` in the `Parameter` window, which is necessary to set parameter values to the original value.

- Save data

`File → All info → name file branch3allinfo.dat`

- Repeat procedure to compute periodic cycles of period four, five, six and seven.
  1. in the XPP window: `F → G → choose set p# → I → G → I → L → I → L ...`
  2. in the XPP window: `U → O → specify n_out # → <Return> → <ESC>`
  3. in the AUTO window: `R → hit ABORT`
  4. in the XPP window:
    - (a) open the `Data` window → scroll down so that the second line is the first line of the window → press `Get`
    - (b) copy the value of `Rzero` from the `Data` window
    - (c) open the `Param` window → set the value of  $\mathcal{R}_0$  as in (b) → hit `Ok`
  5. in the AUTO window: `Run → No` to destroy diagram → `ABORT` to stop integration
  6. repeat steps 4-5 for each branch of the orbit
  7. in AUTO: `F → A → save file branch#allinfo.dat → Ok`

8. in AUTO:  $F \rightarrow R \rightarrow YES$

## 5.4 Plotting the AUTO bifurcation diagram

Each AUTO output file (`branchPallinfo.dat`, where  $P$  is the period) contains the 13 columns named below. A few comments:

**ptype** The point type is an integer between 1 and 4. For our purposes it is always either 1 (stable fixed point) or 2 (unstable fixed point), but in other applications it could be 3 (stable periodic orbit) or 4 (unstable periodic orbit).

**branch** The branch number indicates the distinct component of the branch on which the given state lies. For example, a three-cycle will have three branches and the branch number will be 1, 2 or 3.

**period** The period is always listed as 0 for a discrete map. It is non-zero (and meaningful) for periodic orbits of continuous-time systems.

See Ermentrout's book [3, §9.2.1, p.222].

```
auto.colnames <- c(
  "ptype",    # point type (1=stable, 2=unstable)
  "branch",   # branch number
  "R0",       # first active parameter
  "alpha",    # second active parameter
  "period",   # (not meaningful for discrete maps)
  "shi",      # max S coordinate value
  "ihi",      # max I coordinate value
  "slo",      # min S coordinate value
  "ilo",      # min I coordinate value
  "ev1.re",   # real part of first eigenvalue
  "ev1.im",   # imaginary part of first eigenvalue
  "ev2.re",   # real part of second eigenvalue
  "ev2.im"    # imaginary part of second eigenvalue
)
```

We read all the data from the AUTO output files into a list of data frames. We sort the unique `periods` vector so the index `iper` is also the period (this will be true in general only if all periods occur from 1 to the number of periods).

```

period.set <-
  sort(unique.periods[unique.periods <= maxper])
auto.data <- lapply(period.set,
  function(iper)
    read.table(
      paste("AUTO/branch",iper,"allinfo.dat", sep=""),
      col.names=auto.colnames))

```

We now plot each branch of the AUTO bifurcation diagram.

```

autoplot <- function(auto.data, period.set,
  xlim=c(0,33), ylim=c(-8,-2), ... ) {
  ## create empty plot with no y-axis annotation
  plot(x=0, type="n", xaxs="i", yaxt="n",
    xlim=xlim, ylim=ylim, cex.axis=1.3, cex.lab=1.5,
    xlab=expression(
      paste("Basic Reproduction Number, ", R[0])),
    ylab="Prevalence I/N", ...)
  ## add y-axis annotation
  y.ticks <- 0:-8
  y.label <- paste("10", y.ticks, sep="^")
  axis(2, at=y.ticks, las=2, label=parse(text=y.label))
  ## plot the branch associated with each period
  for (iper in period.set) {
    with(auto.data[[iper]],{
      stable.pts <- which(ptype==1)
      unstable.pts <- which(ptype==2)
      points(R0[unstable.pts],ihi[unstable.pts],
        col=iper,pch=".")
      points(R0[stable.pts],ihi[stable.pts],
        col=iper,pch=19, cex=0.4)
      ## add black dot inside yellow (col=7) points:
      if (iper==7)
        points(R0[stable.pts],ihi[stable.pts],pch=".")
    })
  }
  abline(v=17, col="red", lty="dashed")
}

```

```
autoplot( auto.data, period.set )
```

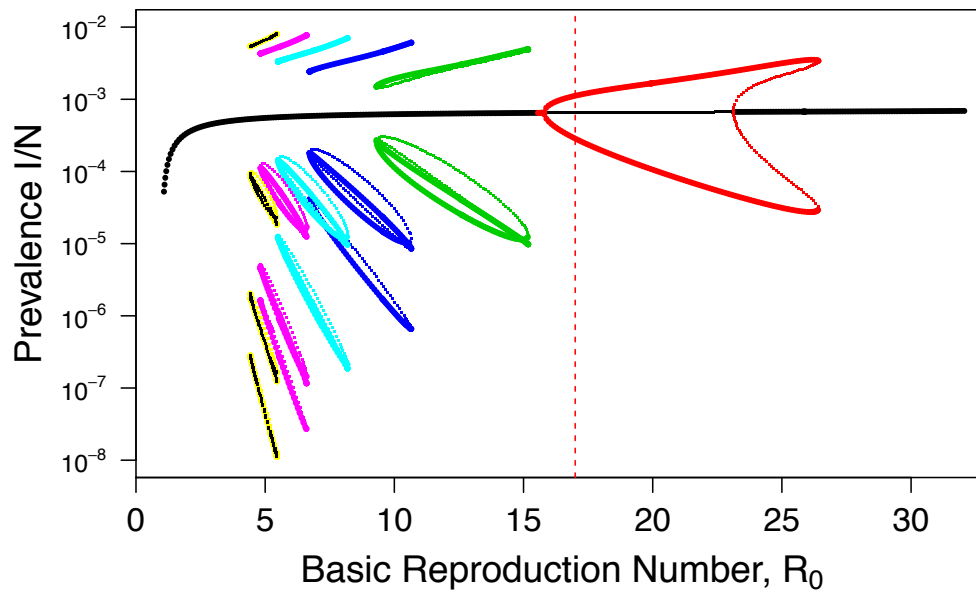


Figure 3: AUTO bifurcation diagram for the sinusoidally forced SIR model (1) with mean infectious period  $1/\gamma = 13$  days, mean lifetime  $1/\mu = 50$  years and amplitude of seasonality  $\alpha = 0.08$ .  $\mathcal{R}_0 = 17$ , appropriate for measles, is highlighted by a dashed vertical line. Compare figure 2.

**Run time for this knitr document.** The start and end times, respectively, were:

```
## [1] "2013-04-14 21:29:35 EDT"  
## [1] "2013-04-14 21:32:27 EDT"
```

## References

- 1 Krylova O. Predicting epidemiological transitions in infectious disease dynamics: Smallpox in historic London (1664-1930) [PhD]. McMaster University, Canada; 2011.
- 2 Maystruk V. XPPAUT to Produce Bifurcation Diagram as Demonstrated with Seasonally Forced SIR and SEIR Models; 2006.
- 3 Ermentrout B. Simulating, analyzing, and animating dynamical systems: a guide to XPPAUT for researchers and students. Software, Environments, and Tools. Philadelphia: Society for Industrial and Applied Mathematics; 2002.
- 4 Doedel E. AUTO: software for continuation and bifurcation problems in ordinary differential equations; 2007. Available from: <http://indy.cs.concordia.ca/auto>.
- 5 Kuznetsov YA. Elements of applied bifurcation theory. vol. 112 of Applied Mathematical Sciences. 3rd ed. New York: Springer-Verlag; 2004.
- 6 Wiggins S. Introduction to applied nonlinear dynamical systems and chaos. vol. 2 of Texts in applied mathematics. 2nd ed. New York: Springer-Verlag; 2003.