

## Stochastic fluctuations in the susceptible-infective-recovered model with distributed infectious periods

Andrew J. Black and Alan J. McKane

*Theory Group, School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, United Kingdom*

Ana Nunes and Andrea Parisi

*Departamento de Física and Centro de Física Teórica e Computacional, Faculdade de Ciências, Universidade de Lisboa, P-1649-003 Lisboa Codex, Portugal*

(Received 3 February 2009; published 19 August 2009)

We investigate a stochastic model of infection dynamics based on the Susceptible-Infective-Recovered (SIR) model, where the distribution of the recovery times can be tuned, interpolating between exponentially distributed recovery times, as in the standard SIR model, and recovery after a fixed infectious period. This is achieved by introducing  $L$  infective classes, as compared to 1 in the standard model. For large populations, the spectrum of fluctuations around the deterministic limit of the model can be computed analytically. The demographic stochasticity has the effect of transforming the decaying oscillations of the deterministic model into sustained oscillations in the stochastic formulation. We find that the amplification of these stochastic oscillations increases with  $L$ , as well as their coherence in frequency. For large values of  $L$  (of the order of 10 and greater), the height and position of the peak of the power spectra changes little and is described well by the model with fixed recovery period ( $L \rightarrow \infty$ ). In this limit we give a closed-form expression for the power spectrum of fluctuations of infective individuals.

DOI: [10.1103/PhysRevE.80.021922](https://doi.org/10.1103/PhysRevE.80.021922)

PACS number(s): 87.10.Mn, 02.50.Ey, 05.40.-a

### I. INTRODUCTION

There are ongoing debates in the field of population dynamics regarding the driving mechanisms of the noisy oscillations found in ecological and epidemiological data [1] and the different approaches to modeling them [2]. This mainly centers around the relative importance of deterministic versus stochastic forces on the dynamics of the system. In epidemiology, where detailed data is more abundant, demographic stochasticity has been shown to give rise to recurrent epidemic outbreaks, which cannot be related to external forcing [3,4].

A stochastic theory developed for a predator-prey model [5] and then applied to a Susceptible-Infective-Recovered (SIR) model with births and immigration [6] helped elucidate the mechanism behind this phenomenon. The power spectrum of the infective time series was derived and shown to be determined, both in the presence and in the absence of external forcing, by the resonance of internal noise with a frequency which could be calculated from the model. For the parameter values that correspond to endemic diseases (e.g., measles, rubella, and whooping cough), the amplitude of the resonant fluctuations are comparable, even in large systems, to the oscillations induced by seasonal forcing.

Within a deterministic framework, several modifications of the basic SIR dynamics have been explored with a view to obtaining robust unforced oscillations as well as more biologically realistic models. Some examples that can be found in the mathematical epidemiology literature are higher-order nonlinearities in the infection term [7,8], age-structured populations [9], delays [10], and coupling with pathogen evolution [11]. More recently, it has been shown that if the network of contacts in the population is taken to evolve as a consequence of disease awareness, the deterministic descrip-

tion of Susceptible-Infected-Susceptible (SIS), SIR and Susceptible-Infected-Recovered-Susceptible (SIRS) dynamics in the uncorrelated pair approximation has an oscillatory phase in a small region of parameter space [12–14].

The coloring of the basic model with a virtually endless palette of additional complexities does of course produce sustained oscillations, and more complex behavior. However, none of these modifications predicts the regular patterns of recurrent epidemics found in many data sets for a significant range of realistic parameter values [15].

Discrete SIR type models inspired by the dynamics of excitable media have been much more successful in producing robust unforced oscillations, provided that there is enough mixing in the interactions [16–21]. In contrast with the continuous models in mathematical epidemiology, which assume constant rates of recovery and immunity waning, these discrete models assume that both recovery and immunity loss occur at a fixed number of time steps after infection. In the context of loss of immunity, this is a very unrealistic assumption, and the oscillatory behavior found in this approach must be taken as an artifact of the model.

For many diseases the infectious period is well defined and the standard assumption of a constant recovery rate, and thus exponentially distributed recovery times, is epidemiologically unrealistic [22–24]. The change to more realistic recovery profiles has been shown to have limited consequences for the deterministic system [25–27], but for the stochastic version of the model, less dispersed infectious periods destabilize it, leading to larger stochastic fluctuations around the endemic equilibrium [28].

In [29] the power spectrum of a stochastic SIR model with a fixed infectious period was numerically computed. For a large parameter range, the amplitude and coherence (power centered about the peak of the spectrum) of the fluctuation

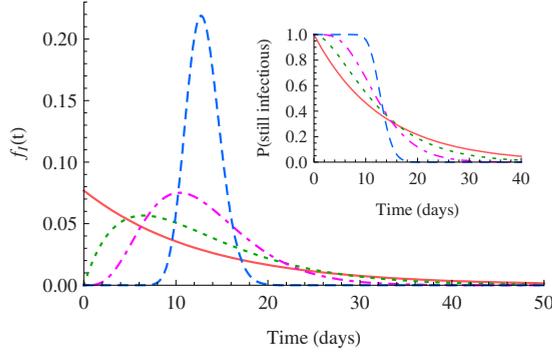


FIG. 1. (Color online) (Main) The distribution of infectious periods,  $f_i(t)$ , for  $L=1, 2, 5$ , and  $50$  (solid, dotted, dot-dashed, and dashed lines, respectively). The case  $L=1$  corresponds to the exponential distribution of the standard SIR model. The inset shows the probability that an individual is still infectious at time  $t$ : for large  $L$  it approaches a step function, where all individuals remain infectious for a constant period of time. The average infectious period,  $1/\gamma$ , is 13 days.

tuations were found to be enhanced with respect to the standard SIR model [6]. In this paper, we establish these results analytically and show that the amplification of demographic stochasticity is large enough for the behavior of moderately sized systems to be akin to the self-sustained oscillations typical of the discrete versions of the model.

## II. MODEL AND ANALYSIS

In order to obtain infectious period distributions that interpolate between exponentially distributed and constant infectious periods, we follow Lloyd [27] and split up the recovery process into a number of stages in which the time spent in each class is exponentially distributed [30]. When susceptible individuals are infected they enter the first infective class, from which they move with a constant rate to the next class, and so forth until recovery. If  $L$  is the total number of infective classes, by fixing the transition rate from class  $I_j$  to class  $I_{j+1}$  equal to  $1/L\gamma$ , the total average recovery period is fixed to  $1/\gamma$ . This produces gamma-distributed recovery profiles, with the constant recovery profile obtained in the limit of large number of infectious classes  $L$  (see Fig. 1). The parameter  $L$  can effectively be fixed in the model by fitting the infectious period distribution to data [22,24]. Previous authors have tended to use smaller values (less than 20) to achieve good fits to data [31–33].

The model has equal birth and death rates  $\mu$  so that the total population size  $N$  is kept constant. The processes through which the system evolves are thus

$$\text{Infection } S + I_j \xrightarrow{\beta} I_j + I_1 \quad j = 1, \dots, L,$$

$$\text{Birth/death } \{I_j, R\} \xrightarrow{\mu} S \quad j = 1, \dots, L,$$

$$\text{Recovery } I_j \xrightarrow{L\gamma} I_{j+1} \quad j = 1, \dots, L-1,$$

$$I_L \xrightarrow{L\gamma} R, \quad (1)$$

where  $S$  is the total number of susceptibles,  $I = \sum_{j=1}^L I_j$  is the total number of infectives,  $I_j$  are the infectives in class  $j$ , and  $\beta$  is the contact rate among individuals. Since  $N = S + I + R$  is constant, we eliminate the number of recovered from our equations by using  $R = N - S - I$ .

There are two ways in which the model may be investigated: For finite  $L$  it can be simulated with Gillespie's algorithm [34] (or one of the more efficient methods based on it [35]), and it can be analyzed analytically by constructing a master equation corresponding to processes (1) and performing van Kampen's system-size expansion [5,6,36] on this equation. There are  $L+1$  independent variables which describe the state of the system, but as our interest is on the total number of infectives, the variables we use to describe the system are  $\sigma \equiv \{S, I, I_2, \dots, I_L\}$ , where  $I_1$  is given in terms of the other variables,  $I_1 = I - \sum_{j=2}^L I_j$ . The master equation then takes the form

$$\frac{dP(\sigma; t)}{dt} = \sum_{\sigma' \neq \sigma} T(\sigma|\sigma')P(\sigma'; t) - \sum_{\sigma' \neq \sigma} T(\sigma'|\sigma)P(\sigma; t), \quad (2)$$

where  $T(\sigma|\sigma')$  is the transition rate from the state  $\sigma'$  to the state  $\sigma$ .

From the definition of the model in Eq. (1) we can read off the transitions rates; they are generalizations of those found in [6]. Listed together with the reaction they represent, they are

$$(1) S + I \xrightarrow{\beta} I + I_1$$

$$T(S-1, I+1, I_2, \dots | S, \mathbf{I}) = \frac{\beta SI}{N},$$

$$(2) I_1 \xrightarrow{\mu} S$$

$$T(S+1, I-1, I_2, \dots | S, \mathbf{I}) = \mu \left( I - \sum_{j=2}^L I_j \right),$$

$$(3) I_j \xrightarrow{\mu} S$$

$$T(S+1, I-1, \dots, I_j-1, \dots | S, \mathbf{I}) = \mu I_j,$$

$$(4) R \xrightarrow{\mu} S$$

$$T(S+1, I, \dots | S, \mathbf{I}) = \mu(N - S - I),$$

$$(5) I_1 \xrightarrow{L\gamma} I_2$$

$$T(S, I, I_2+1, \dots | S, \mathbf{I}) = L\gamma \left( I - \sum_j I_j \right),$$

$$(6) I_j \xrightarrow{L\gamma} I_{j+1}$$

$$T(S, I, \dots, I_j-1, I_{j+1}+1, \dots | S, \mathbf{I}) = L\gamma I_j,$$

$$(7) I_L \xrightarrow{L\gamma} R$$

$$T(S, I-1, I_2, \dots, I_L-1 | S, \mathbf{I}) = L\gamma I_L,$$

where  $\mathbf{I}=(I, I_2, \dots, I_L)$ .

We may now apply the system-size expansion. At leading order in the system size  $N$ , this expansion gives the mean-field model, while higher-order terms describe the stochastic fluctuations. One starts by introducing the new variables:  $S = \phi N + x_0 \sqrt{N}$ ,  $I = \psi N + x_1 \sqrt{N}$ , and  $I_j = \psi_j N + x_j \sqrt{N}$  with  $j \geq 2$ , where  $\phi$  is the fraction of susceptibles,  $\psi$  is the fraction of total infectives, and  $\psi_j$  is the fraction of infectives in class  $j$ , with  $x_0, \dots, x_L$  describing the stochastic corrections. The full details involved in carrying out the expansion are given in Appendix A. To leading order, we find the mean-field (deterministic) equations for the fractions, which, introducing the new parameters  $\hat{\beta} = \beta/\gamma$  and  $\hat{\mu} = \mu/\gamma$ , and moving to a scaled time  $\tau = \gamma t$ , have the form

$$\dot{\phi} = -\hat{\beta}\phi\psi + \hat{\mu}(1-\phi),$$

$$\dot{\psi} = \hat{\beta}\phi\psi - \hat{\mu}\psi - L\psi_L,$$

$$\dot{\psi}_2 = -(L + \hat{\mu})\psi_2 + L\left(\psi - \sum_{j=2}^L \psi_j\right),$$

$$\dot{\psi}_j = -(L + \hat{\mu})\psi_j + L\psi_{j-1}, \quad j = 3, \dots, L. \quad (3)$$

Apart from a trivial solution ( $\psi^* = 0$ ;  $\psi_j^* = 0$ ,  $j = 2, \dots, L$ ), there is a unique fixed point for Eqs. (3). The full form is given in Appendix A, but keeping only terms of order  $\hat{\mu}$ , it is  $\phi^* = [1 + (1+1/L)\hat{\mu}/2]/\hat{\beta}$ ,  $\psi^* = (1-1/\hat{\beta})\hat{\mu}$ , and  $\psi_j^* = (1-1/\hat{\beta})\hat{\mu}/L$ .

The variables  $x_0, \dots, x_L$  describe the fluctuations around a trajectory of the mean-field system. We will be especially interested in the fluctuations when the transients of the deterministic equations have died out and the system is in equilibrium, fluctuating about the fixed point. The fluctuations obey a linear Fokker-Planck equation, which is equivalent to a set of Langevin equations of the form [5,6,36]

$$\frac{dx_i}{d\tau} = \sum_{j=0}^L A_{ij}x_j + \eta_i(\tau), \quad (4)$$

where  $\eta_i(\tau)$  are Gaussian noise terms with zero mean and satisfying  $\langle \eta_i(\tau)\eta_j(\tau') \rangle = B_{ij}\delta(\tau-\tau')$ . Therefore the Gaussian stochastic process describing the fluctuations is completely characterized by two  $L+1$  dimensional matrices,  $A_{ij}$  and  $B_{ij}$ . These are the final result of carrying out the expansion and depend on the time  $\tau$  through the solutions of the mean-field equations  $\phi$ ,  $\psi$ , and  $\psi_j$ . As mentioned above we will be interested in evaluating these at the fixed point, and as a consequence the matrices  $A$  and  $B$  cease to be time dependent. Explicit expressions for these two matrices are given in Appendix A.

To search for sustained oscillations we need to find the power spectrum for the total number of infectives,  $P_L(\omega) = \langle |\tilde{x}_1|^2 \rangle$ , where  $\tilde{x}_1$  is the Fourier transform of  $x_1(\tau)$ , the subscript  $L$  denotes that the whole procedure depends on the number of infective classes. By taking the Fourier transform of Eq. (4) we obtain

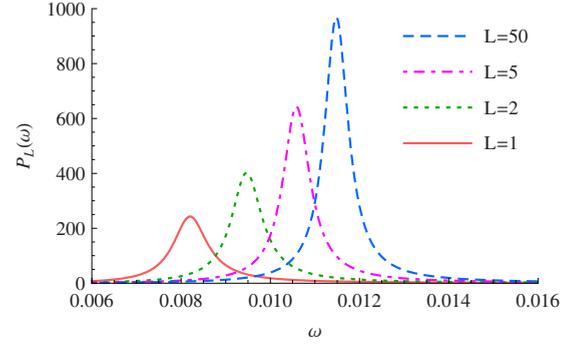


FIG. 2. (Color online) Theoretical power spectra for the fluctuations of the total number of infectives when  $L=1, 2, 5$ , and  $50$ . The case  $L=1$  corresponds to exponential recovery. For increasing  $L$ , there is a shift of the resonant frequency and an increase in the amplitude of fluctuations. Other parameters are  $\beta=1.3$ ,  $1/\gamma=13$ , and  $\mu=5.5 \times 10^{-5}$ .

$$\sum_{j=0}^L S_{ij}\tilde{x}_j + \tilde{\eta}_i = 0, \quad (5)$$

where  $S_{ij} = A_{ij} - i\omega\delta_{ij}$ . The power spectrum is then obtained as

$$P_L(\omega) = \frac{\sum_{ij} B_{ij} C_i(\omega) C_j^*(\omega)}{D(\omega) D^*(\omega)}, \quad (6)$$

where  $C_i(\omega)$  is the co-factor of the matrix  $S$  in row  $i$  and column  $j=1$ , and  $D(\omega)$  is the determinant of the full  $S$  matrix. In practice the matrices are computed numerically, and  $P_L(\omega)$  is computed from these using a symbolic package such as MATHEMATICA [37].

In general  $A_{ij}$  will have  $L+1$  eigenvalues. Since  $S_{ij} = A_{ij} - i\omega\delta_{ij}$ , the zeros of the determinant of  $S$  in  $\omega$  space will be at positions which are  $-i$  multiplied by the eigenvalues. At these values, the power spectrum will have a pole in the complex  $\omega$  plane. Therefore if there are  $\ell$  complex eigenvalues of  $A$ , we would expect to find  $\ell$  peaks in the spectrum at frequencies which are approximately equal to the imaginary parts of the eigenvalues. In numerical simulations we only ever see one peak (see Fig. 4), so the question arises: is there only ever one set of complex eigenvalues? Numerical determination of the eigenvalues of  $A$  for small  $L$  shows that there are typically more than one pair of complex eigenvalues, but that there is always one dominant pair, with a real part several orders of magnitude smaller than the others. This implies that this dominant eigenvalue is very close to the real  $\omega$  axis and so will give a large peak and therefore large amplitude fluctuations. The imaginary part of the dominant eigenvalue is also orders of magnitude smaller than those of the others, and therefore the tiny peaks resulting from these other complex eigenvalues will be at much larger frequencies.

The theoretical power spectra resulting from the infectious period distributions in Fig. 1 are presented in Fig. 2. The case previously considered [6] corresponds to  $L=1$ , and it is clear that both the enhancement of the amplitude and the coherence of the power spectrum increases as  $L$  increases. The change is more pronounced for low values of  $L$  follow-

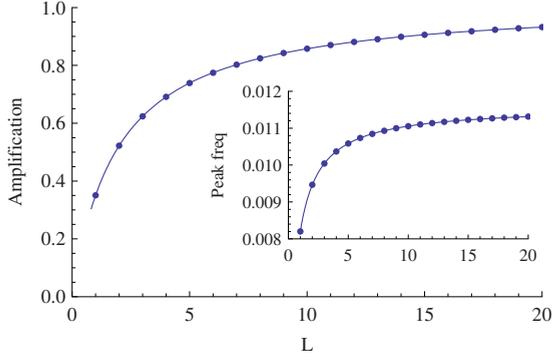


FIG. 3. (Color online) (Main) Increase in amplification of  $P_L(\omega)$  with increasing  $L$ , where amplification is defined as the area under the power spectrum, which is equal to the mean squared variance of the time series  $x_1$ . (Inset) Drift of the peak position of  $P_L(\omega)$  for increasing  $L$ . In both cases  $\beta=1.3$ ,  $1/\gamma=13$ , and  $\mu=5.5 \times 10^{-5}$ . Both curves are perfectly fitted by an expression of the kind:  $a - b/(c+L)$  (continuous line), which shows that the underlying dependence on  $L$  is simple and that the power spectrum converges to a definite shape as  $L \rightarrow \infty$ .

ing the large changes in the recovery profile at those low values. Therefore for the values of  $L$  which are usually thought to be applicable ( $L$  of the order of 10 or 20 [31,32]), an enhancement of three or four times that found with exponential recovery ( $L=1$ ) can be seen. The position of the peak (given by the fixed points of the mean-field equations) also shifts with  $L$ . Figure 3 is a plot of the amplification (which is proportional to the amplitude [6]) and position of the peak (endogenous frequency) of the power spectrum as a function of  $L$ . The change in these quantities with  $L$  is smooth and appears to approach a fixed limit for large  $L$ . They are well fitted by an expression of the form  $a - b/(c+L)$ , where  $a$ ,  $b$ , and  $c$  are constants, reinforcing the notion that they have a finite limit as  $L \rightarrow \infty$ .

Since the position and peak of the power spectrum change little for values of  $L$  of the order of 10 or above, it is gratifying to find that the power spectrum for  $L \rightarrow \infty$  can be obtained in closed form, at least for small  $\hat{\mu}$ . The details of the calculation are given in Appendix B, but we will give the final result here. We first define  $C_j(\omega) = (-1)^{L-1} L^L \hat{C}_j(\omega)$  and  $D(\omega) = (-1)^{L-1} L^L \hat{D}(\omega)$ , so that power spectrum (6) becomes

$$P_L(\omega) = \frac{\sum_{i=0}^L \sum_{j=0}^L \hat{C}_i(\omega) B_{ij} \hat{C}_j^*(\omega)}{\hat{D}(\omega) \hat{D}^*(\omega)}. \quad (7)$$

Now it turns out that  $\hat{C}_j(\omega)$  and  $\hat{D}(\omega)$  have a finite limit as  $L \rightarrow \infty$ . Specifically, as shown in Appendix B the numerator of Eq. (7) is

$$\begin{aligned} \sum_{i=0}^L \sum_{j=0}^L \hat{C}_i(\omega) B_{ij} \hat{C}_j^*(\omega) &= \hat{\mu} \left( 1 - \frac{1}{\hat{\beta}} \right) (\hat{\mu}^2 + \omega^2)^{-1} \\ &\times [\hat{\mu}^2 (\hat{\beta}^2 - 2\hat{\beta} + 2) + \omega^2] \\ &\times (e^{2\hat{\mu}} - 2e^{\hat{\mu}} \cos \omega + 1), \end{aligned} \quad (8)$$

and the denominator is

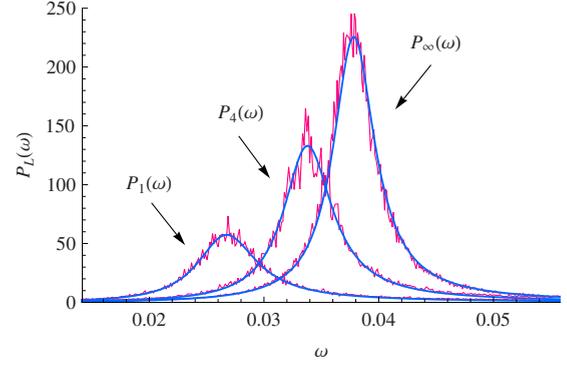


FIG. 4. (Color online) Comparison between the analytic power spectra (solid blue lines) given by Eq. (6) and numerical simulations of the SIR model (noisy red lines) for  $L=1, 4$ , and  $\infty$ , for which there is perfect agreement. Other parameter values are  $\beta=1.32$ ,  $1/\gamma=8$ ,  $\mu=6 \times 10^{-4}$ , and a population size of  $N=10^6$  individuals. The numerical curves were obtained by averaging the power spectra of 200 realizations for each of the three cases considered.

$$\begin{aligned} \hat{D}(\omega) \hat{D}^*(\omega) &= e^{2\hat{\mu}} \left[ \left( \hat{\beta} \hat{\mu} - 1 - \frac{\hat{\mu}}{2} \right)^2 + \omega^2 \right] + e^{\hat{\mu}} (2 + \hat{\mu}) \\ &\times \left[ \cos \omega \left( \hat{\beta} \hat{\mu} - 1 - \frac{\hat{\mu}}{2} \right) - \omega \sin \omega \right] \\ &+ \left( 1 + \frac{\hat{\mu}}{2} \right)^2. \end{aligned} \quad (9)$$

A numerical determination of the zeros of function (9) shows that there is a single pair of complex roots with small real and imaginary parts, which give the dominant peak in the spectrum. The analytically derived spectra are compared to those obtained numerically in Fig. 4. The limit  $L \rightarrow \infty$  corresponds to the case where recovery occurs at a fixed period of time after infection. The power spectrum for this case was obtained by simulating a standard SIR model (with a single infective class) using a fixed time step. The stochastic infection and birth/death processes are implemented as usual, but with recovery occurring at a constant number of time steps after infection [38]. The numerical spectra have intentionally been left relatively noisy to help distinguish them from the analytic results, otherwise the agreement is excellent.

### III. DISCUSSION AND CONCLUSION

There are a wide range of models of infection dynamics, from the very simple deterministic SIR model right through to agent-based models which incorporate details of the individuals in the community under investigation. Simple deterministic models have the advantage of being able to be understood analytically, but the greater the degree of realism the more difficult this becomes. In the past the introduction of stochasticity has been an example of this: while stochastic effects are undoubtedly important in many situations, their inclusion meant a resort to numerical methods rather than analysis. However recently, by using techniques originally developed in nonequilibrium statistical physics [5,6,36], the

stochastic SIR model has been opened up to analytic investigation.

In this paper we have moved one step further and shown that the SIR model with  $L$  infectious stages, which has a much more realistic recovery profile, can be analyzed almost as simply as the standard SIR model (which has  $L=1$ ). The sustained and amplified oscillations found in the standard SIR model [6] are even more evident for  $L>1$ , with the frequency of the oscillations and their amplitude increasing with  $L$ . Typical values of  $L$  estimated from data lie between 10 and 20 [31,32], and for these values the frequency and amplitude of the oscillations are already near to the asymptotic limit  $L\rightarrow\infty$ . In this limit the power spectrum may be obtained in closed form, verifying the small changes that occur in the nature of the spectrum for large  $L$ .

Some aspects of this staged SIR model have been investigated previously. Grossman [26] studied the deterministic SIR model incorporating a fixed infectious period. Later this work was elaborated by Lloyd [27,28] who expanded the results to include gamma-distributed infectious periods, as utilized in this paper, and studied the stochastic version of the model numerically. He found that the fixed points became less stable with increasing  $L$ , which he interpreted as a “destabilization” of the SIR model. Our result that the total amplification (which is proportional to the mean variance of the time series) increases with  $L$  is consistent with Lloyd’s earlier result that the damping time of the deterministic system increased with  $L$ . However we should stress that the frequency of the damped oscillations of the deterministic system is only approximately the same as the frequency of the sustained oscillations, due to the additional frequency dependence in the numerator of power spectrum (6).

While destabilization has been discussed previously, the increase in the endogenous frequency of the system with increasing  $L$  has received comparatively little attention. For example, the exponentially distributed model, parametrized for measles, predicts a natural period of oscillation of 2 years, whereas the fixed infectious period version ( $L\rightarrow\infty$ ) predicts 1.5 years. We have not included seasonal forcing in our analysis, but we would expect that a system with a higher endogenous frequency would be more unstable to seasonal forcing. This would be in line with the findings of previous authors [28,31], in the context of deterministic SIR models.

There are several other extensions of the work carried out in this paper which we believe would be interesting and fruitful. External immigration has not been included in the present study. Since immigration has been shown to have a stabilizing effect on oscillations in the standard SIR model [39], it would be of interest to fully quantify this in a staged SIR model. Additionally, in order to compare our results effectively with data, we need to analyze the SEIR version of model, so as to capture the most realistic recovery profile. We believe that many of the studies can be carried out using similar techniques to those utilized in this paper, and so still further extend the range of models of infectious diseases which can be studied analytically.

#### ACKNOWLEDGMENTS

A.J.B. wishes to thank the EPSRC (U.K.) for the award of

a postgraduate grant. A.N. gratefully acknowledges financial support from the Foundation of the University of Lisbon and the Portuguese Foundation for Science and Technology (FCT) under Contract No. POCTI/ISFL/2/618. A.P. acknowledges funding from the European Commission under Grant No. MEXT-CT-2004-14338.

#### APPENDIX A: APPLYING THE SYSTEM-SIZE EXPANSION

In this appendix we give details of the application of the van Kampen system-size expansion, and also give some of the results (such as the form of the matrices  $A$  and  $B$ ) that are too unwieldy to include in the main text.

The master equation for the time evolution of the probability distribution function is given in Eq. (2). To write it out explicitly, as we need to in order to apply the system-size expansion, we introduce the step operators [5,6,36]

$$\epsilon_s^{\pm 1} f(S, I, I_2, \dots, I_L) = f(S \pm 1, I, I_2, \dots, I_L),$$

$$\epsilon_1^{\pm 1} f(S, I, I_2, \dots, I_L) = f(S, I \pm 1, I_2, \dots, I_L),$$

$$\epsilon_j^{\pm 1} f(S, I, I_2, \dots, I_L) = f(S, I, I_2, \dots, I_j \pm 1, \dots, I_L).$$

Then Eq. (2) with transition rates (1) becomes

$$\begin{aligned} \frac{dP(\sigma; t)}{dt} = & \left\{ (\epsilon_s^{+1} \epsilon_1^{-1} - 1) \frac{\beta SI}{N} + (\epsilon_s^{-1} \epsilon_1^{+1} - 1) \mu \left( I - \sum_{j=2}^L I_j \right) \right. \\ & + \sum_{j=2}^L (\epsilon_s^{-1} \epsilon_1^{+1} \epsilon_j^{+1} - 1) \mu I_j + (\epsilon_s^{-1} - 1) \mu (N - S - I) \\ & + (\epsilon_2^{-1} - 1) \gamma \left( I - \sum_{j=2}^L I_j \right) + \sum_{j=2}^{L-1} (\epsilon_j^{+1} \epsilon_{j+1}^{-1} - 1) \gamma I_j \\ & \left. + (\epsilon_1^{+1} \epsilon_L^{+1} - 1) \gamma I_L \right\} P(\sigma; t). \end{aligned} \quad (\text{A1})$$

Making the change of variables

$$S = N\phi + \sqrt{N}x_0,$$

$$I = N\psi + \sqrt{N}x_1,$$

$$I_j = N\psi_j + \sqrt{N}x_j,$$

mentioned in the main text, the step operators may be expanded in a power series in  $N^{-1/2}$ ,

$$\epsilon_s^{\pm 1} = 1 \pm \frac{1}{\sqrt{N}} \frac{\partial}{\partial x_0} + \frac{1}{2N} \frac{\partial^2}{\partial x_0^2},$$

$$\epsilon_1^{\pm 1} = 1 \pm \frac{1}{\sqrt{N}} \frac{\partial}{\partial x_1} + \frac{1}{2N} \frac{\partial^2}{\partial x_1^2},$$

$$\epsilon_j^{\pm 1} = 1 \pm \frac{1}{\sqrt{N}} \frac{\partial}{\partial x_j} + \frac{1}{2N} \frac{\partial^2}{\partial x_j^2}.$$

Substituting these into the master Eq. (A1) we may identify a hierarchy of equations multiplied by different powers of

$N^{-1/2}$ . At leading order we obtain Eqs. (3). The unique non-trivial fixed point of these equations is given by

$$\begin{aligned}\phi^* &= \frac{\hat{\mu}\hat{\beta}}{1 - \left(1 + \frac{\hat{\mu}}{L}\right)^{-L}}, \\ \psi_j^* &= \left(1 - \frac{\hat{\mu}}{\hat{\beta}}\right) - \left(1 + \frac{\hat{\mu}}{L}\right)^{-L}, \\ \psi_j^* &= \frac{\frac{\hat{\mu}}{L}\left(1 + \frac{\hat{\mu}}{L}\right)^{-j}}{\left(1 + \frac{\hat{\mu}}{L}\right)^L - 1} \left[ \left(1 - \frac{\hat{\mu}}{\hat{\beta}}\right) \left(1 + \frac{\hat{\mu}}{L}\right)^L - 1 \right], \\ j &= 2, \dots, L.\end{aligned}\quad (\text{A2})$$

The form of these solutions keeping only terms of order  $\hat{\mu}$  is

$$\phi^* = \frac{1}{\hat{\beta}} \left[ 1 + \frac{\hat{\mu}}{2} \left(1 + \frac{1}{L}\right) \right] = \frac{1}{\hat{\beta}} [1 + \hat{\mu}X(L)],$$

$$\psi^* = \hat{\mu} \left(1 - \frac{1}{\hat{\beta}}\right),$$

$$\psi_j^* = \frac{\hat{\mu}}{L} \left(1 - \frac{1}{\hat{\beta}}\right), \quad j = 2, \dots, L, \quad (\text{A3})$$

as given in the main text.

At next-to-leading order in the expansion of the master equation, one finds a linear Fokker-Planck equation for the fluctuation variables  $x_0, \dots, x_L$  of the following type:

$$\frac{\partial \Pi}{\partial \tau} = - \sum_{i=0}^L \sum_{j=0}^L A_{ij} \frac{\partial [x_j \Pi]}{\partial x_i} + \frac{1}{2} \sum_{i=0}^L \sum_{j=0}^L B_{ij} \frac{\partial^2 \Pi}{\partial x_i \partial x_j}, \quad (\text{A4})$$

where the two matrices  $A_{ij}$  and  $B_{ij}$  which come out of the expansion depend on the time  $\tau$  through  $\phi$ ,  $\psi_0$ , and  $\psi_j$ . Since we are interested in fluctuations about the equilibrium state, both matrices are evaluated at the fixed point [Eq. (A3)]. The Fokker-Planck Eq. (A4) is equivalent to the Langevin Eq. (4) [40,41], and so a complete description of the fluctuations is given by the matrices  $A$  and  $B$ . The explicit forms of  $A_{ij}$  and  $B_{ij}$  evaluated at the fixed point are found to be

$$\begin{aligned}A_{ij} &= \begin{bmatrix} -\hat{\beta}\psi - \hat{\mu} & -\hat{\beta}\phi & 0 & 0 & 0 & \dots & 0 \\ \hat{\beta}\psi & \hat{\beta}\phi - \hat{\mu} & 0 & 0 & 0 & \dots & -L \\ 0 & L & -(2L + \hat{\mu}) & -L & -L & \dots & -L \\ 0 & 0 & L & -(L + \hat{\mu}) & 0 & \dots & 0 \\ 0 & 0 & 0 & L & -(L + \hat{\mu}) & \dots & 0 \\ \vdots & & & & & \ddots & \\ & & & & & & \ddots \end{bmatrix}_{\phi=\phi^*, \psi_j=\psi_j^*}, \\ B_{ij} &= \begin{bmatrix} 2\hat{\mu}(1 - \phi) & \hat{\mu}(\phi - \psi_0 - 1) & -\hat{\mu}\psi_2 & -\hat{\mu}\psi_3 & \dots & -\hat{\mu}\psi_L \\ \hat{\mu}(\phi - \psi_0 - 1) & 2\hat{\mu}(1 - \phi) & \hat{\mu}\psi_2 & \hat{\mu}\psi_3 & \dots & (L + \hat{\mu})\psi_L \\ -\hat{\mu}\psi_2 & \hat{\mu}\psi_2 & 2(L + \hat{\mu})\psi_2 & -L\psi_2 & \dots & 0 \\ -\hat{\mu}\psi_3 & \hat{\mu}\psi_3 & -L\psi_2 & 2(L + \hat{\mu})\psi_3 & \dots & 0 \\ \vdots & & & & \ddots & \end{bmatrix}_{\phi=\phi^*, \psi_j=\psi_j^*}.\end{aligned}$$

## APPENDIX B: THE LARGE $L$ limit

We are able to obtain an analytic form for the power spectrum in the limit  $L \rightarrow \infty$ , provided we work to linear order in  $\hat{\mu}$ . In this case the fixed points take the form of Eq. (A3). In this small  $\hat{\mu}$  approximation the matrices  $A$  and  $B$  are

$$A_{ij} \approx \begin{bmatrix} -\hat{\mu}\hat{\beta} & -1 - \hat{\mu}X(L) & 0 & 0 & 0 & \dots & 0 \\ \hat{\mu}(\hat{\beta} - 1) & 1 + \hat{\mu}X(L) - \hat{\mu} & 0 & 0 & 0 & \dots & -L \\ 0 & L & -2L - \hat{\mu} & -L & -L & \dots & -L \\ 0 & 0 & L & -L - \hat{\mu} & 0 & \dots & 0 \\ 0 & 0 & 0 & L & -L - \hat{\mu} & \dots & 0 \\ \vdots & & & & & \ddots & \end{bmatrix}, \quad (\text{B1})$$

$$B_{ij} \approx \begin{bmatrix} 2\theta & -\theta & 0 & 0 & 0 & \dots & 0 \\ -\theta & 2\theta & 0 & 0 & 0 & \dots & \theta \\ 0 & 0 & 2\theta & -\theta & 0 & \dots & 0 \\ 0 & 0 & -\theta & 2\theta & -\theta & \dots & 0 \\ 0 & 0 & 0 & -\theta & 2\theta & \dots & 0 \\ \vdots & & & & & \ddots & \end{bmatrix}, \quad (\text{B2})$$

where in  $B_{ij}$  we have defined  $\theta = \hat{\mu}(1 - \frac{1}{\beta})$ .

To carry out the  $L \rightarrow \infty$  limit, we introduce a sequence of matrices. The matrix  $S_{ij} = A_{ij} - i\omega\delta_{ij}$  is defined in the main text, together with its determinant  $D(\omega)$ . It has  $L+1$  rows and  $L+1$  columns which we label  $0, 1, \dots, L$ . The determinant of the  $(L-1) \times (L-1)$  matrix obtained by omitting the first two rows and first two columns of  $S$ , using approximation (B1) for  $A$ , will be denoted by  $R_{L-1}$ . By expanding the determinant  $D(\omega)$  about the first two columns one finds that

$$D(\omega) = \hat{\mu}(\hat{\beta} - 1)[1 + \hat{\mu}X(L)]R_{L-1}(\omega) - (\hat{\beta}\hat{\mu} + i\omega) \times \{[1 + \hat{\mu}X(L) - (\hat{\mu} + i\omega)]R_{L-1}(\omega) + (-L)^L\}. \quad (\text{B3})$$

It can be seen from Eq. (B1) that the matrix obtained from deleting the first two rows and columns can be defined for an arbitrary value of  $L$ , since the entries for the  $j$ th rows and columns for  $j > 2$  are simply repeated. Therefore we can define  $R_n(\omega)$  for any  $n$ . Expanding the determinant  $R_n(\omega)$  about the first column we find that

$$R_n(\omega) = -(2L + \hat{\mu} + i\omega)(-1)^{n-1}(L + \hat{\mu} + i\omega)^{n-1} - LQ_{n-1}(\omega), \quad (\text{B4})$$

with

$$Q_n(\omega) = \det \begin{bmatrix} -L & -L & -L & \dots \\ L & -L - \hat{\mu} - i\omega & 0 & \dots \\ 0 & L & -L - \hat{\mu} - i\omega & \dots \\ \vdots & & & \ddots \end{bmatrix}.$$

By expanding along the first column we have

$$Q_n(\omega) = -L(-L - \hat{\mu} - i\omega)^{n-1} - LQ_{n-1}(\omega). \quad (\text{B5})$$

For small values of  $n$   $Q_n(\omega)$  can be calculated explicitly. For instance, when  $n=3$  this polynomial is found to be

$$Q_3(\omega) = -L[(L + \hat{\mu} + i\omega)^2 + L(L + \hat{\mu} + i\omega) + L^2].$$

These small- $n$  forms suggest the general result

$$Q_n(\omega) = (-1)^n L \sum_{j=0}^{n-1} L^j (L + \hat{\mu} + i\omega)^{n-1-j}, \quad (\text{B6})$$

which may be proved from Eq. (B5) by induction. This in turn allows us to use Eq. (B4) to find the general expression for  $R_n(\omega)$ ,

$$R_n(\omega) = (-1)^n \sum_{j=0}^n L^j (L + \hat{\mu} + i\omega)^{n-j}. \quad (\text{B7})$$

We wish to isolate the  $L$  dependence in  $R_n(\omega)$ , so that eventually we may take the  $L \rightarrow \infty$  limit. We do this by writing  $R_n(\omega)$  in a different form as follows:

$$\begin{aligned} R_n(\omega) &= (-1)^n \sum_{j=0}^n L^j \left[ \sum_{k=0}^{n-j} \frac{(n-j)! L^{n-j-k}}{k!(n-j-k)!} (\hat{\mu} + i\omega)^k \right] \\ &= (-1)^n L^n \sum_{m=0}^n \sum_{k=0}^m \frac{m!}{k!(m-k)!} \left( \frac{\hat{\mu} + i\omega}{L} \right)^k \\ &= (-1)^n L^n \sum_{k=0}^n \sum_{m=k}^n \frac{m!}{k!(m-k)!} \left( \frac{\hat{\mu} + i\omega}{L} \right)^k \\ &= (-1)^n L^n \sum_{k=0}^n \frac{(n+1)!}{(k+1)!(n-k)!} \left( \frac{\hat{\mu} + i\omega}{L} \right)^k \\ &= (-1)^n L^{n+1} \sum_{k=0}^n \alpha_{n+1}(k) (\hat{\mu} + i\omega)^k, \end{aligned}$$

where we used  $m = n - j$ , and defined

$$\alpha_{n+1}(k) = \frac{1}{L^{k+1}} \frac{(n+1)!}{(k+1)!(n-k)!}. \quad (\text{B8})$$

It is useful to introduce  $\hat{R}_n(\omega)$  through  $R_n(\omega) = (-1)^n L^{n+1} \hat{R}_n(\omega)$ , so that

$$\hat{R}_n(\omega) = \sum_{k=0}^n \alpha_{n+1}(k) (\hat{\mu} + i\omega)^k, \quad (\text{B9})$$

which results in all the dependence on  $L$  being left in the coefficients  $\alpha_{n+1}(k)$ .

To calculate the factor in the numerator of the power spectrum we need to determine the function  $C_i(\omega)$ , which is the co-factor of the matrix  $S$  in row  $i$  and column 1 [note that this is the  $(i+1)$ th row and the second column, since the labeling of rows and columns begins with 0]. Using similar methods to those above one finds that

$$C_0(\omega) = -\hat{\mu}(\hat{\beta} - 1)R_{L-1}(\omega),$$

$$C_1(\omega) = -(\hat{\mu}\hat{\beta} + i\omega)R_{L-1}(\omega),$$

$$C_j(\omega) = (-1)^{L+j-1} L^{L+1-j} (\hat{\mu}\hat{\beta} + i\omega)R_{j-2}(\omega),$$

where in the last line  $j \geq 2$  and  $R_0(\omega) \equiv 1$ . Defining  $C_j(\omega) = (-1)^{L-1} L^L \hat{C}_j(\omega)$  and  $D(\omega) = (-1)^{L-1} L^L \hat{D}(\omega)$ , power spectrum (6) becomes Eq. (7).

Now we can begin to calculate the various factors in Eq. (7), beginning with the denominator. In terms of  $\hat{D}(\omega)$  and  $\hat{R}(\omega)$ , Eq. (B3) reads

$$\hat{D}(\omega) = [(\hat{\beta}\hat{\mu} - \hat{\mu}X(L) - 1) + i\omega](\hat{\mu} + i\omega)\hat{R}_{L-1}(\omega) - (\hat{\beta}\hat{\mu} + i\omega), \quad (\text{B10})$$

but we can find  $\hat{R}_n(\omega)$  from Eq. (B9),

$$\begin{aligned} \hat{R}_n(\omega) &= \sum_{k=0}^n \alpha_{n+1}(k)(\hat{\mu} + i\omega)^k \\ &= \sum_{m=1}^{n+1} \frac{(n+1)!}{L^m m! (n+1-m)!} \frac{(\hat{\mu} + i\omega)^m}{\hat{\mu} + i\omega} \\ &= \frac{1}{\hat{\mu} + i\omega} \left[ \left( 1 + \frac{\hat{\mu} + i\omega}{L} \right)^{n+1} - 1 \right], \end{aligned} \quad (\text{B11})$$

which has a finite limit for  $n=L-1$  and large  $L$ ,

$$\lim_{L \rightarrow \infty} \hat{R}_{L-1}(\omega) = \frac{e^{\hat{\mu} + i\omega} - 1}{\hat{\mu} + i\omega}. \quad (\text{B12})$$

From Eqs. (B10) and (B12) one obtains

$$\hat{D}(\omega) = e^{\hat{\mu} + i\omega} \left[ \left( \hat{\beta}\hat{\mu} - \frac{\hat{\mu}}{2} - 1 \right) + i\omega \right] + \left( 1 + \frac{\hat{\mu}}{2} \right), \quad (\text{B13})$$

and from this Eq. (9) can be obtained.

The numerator can be written as the sum of the following different contributions:

$$\sum_{i=0}^1 \sum_{j=0}^1 B_{i,j} \hat{C}_i(\omega) \hat{C}_j(\omega), \quad (\text{B14})$$

$$B_{1,L} \hat{C}_1(\omega) \hat{C}_L^*(\omega) + B_{L,1} \hat{C}_L(\omega) \hat{C}_1^*(\omega), \quad (\text{B15})$$

$$\sum_{j=2}^L B_{j,j} \hat{C}_j(\omega) \hat{C}_j^*(\omega), \quad (\text{B16})$$

$$\sum_{j=2}^{L-1} [B_{j,j+1} \hat{C}_j(\omega) \hat{C}_{j+1}^*(\omega) + B_{j+1,j} \hat{C}_{j+1}(\omega) \hat{C}_j^*(\omega)]. \quad (\text{B17})$$

The first two contributions, Eqs. (B14) and (B15), are straightforward to evaluate and lead to the following result in the limit  $L \rightarrow \infty$ :

$$\lim_{L \rightarrow \infty} \sum_{i=0}^1 \sum_{j=0}^1 B_{i,j} \hat{C}_i(\omega) \hat{C}_j(\omega) = 2\theta \frac{\hat{\mu}^2 (\hat{\beta}^2 - \hat{\beta} + 1) + \omega^2}{\hat{\mu}^2 + \omega^2} \times (e^{2\hat{\mu}} - 2e^{\hat{\mu}} \cos \omega + 1)$$

and

$$\begin{aligned} &\lim_{L \rightarrow \infty} \{B_{1,L} \hat{C}_1(\omega) \hat{C}_L^*(\omega) + B_{L,1} \hat{C}_L(\omega) \hat{C}_1^*(\omega)\} \\ &= -2\theta \frac{\hat{\beta}^2 \hat{\mu}^2 + \omega^2}{\hat{\mu}^2 + \omega^2} (e^{2\hat{\mu}} - 2e^{\hat{\mu}} \cos \omega + 1). \end{aligned}$$

The third and fourth contributions, Eqs. (B16) and (B17), when taken together equal

$$\theta (\hat{\beta}^2 \hat{\mu}^2 + \omega^2) \left\{ \sum_{n=1}^{L-2} [\hat{R}_n - \hat{R}_{n-1}] [\hat{R}_n^* - \hat{R}_{n-1}^*] + [|\hat{R}_0|^2 + |\hat{R}_{L-2}|^2] \right\}. \quad (\text{B18})$$

Using Eq. (B11) it is found that the sum vanishes for large  $L$ , as does  $|\hat{R}_0|^2$ , with only  $|\hat{R}_{L-2}|^2$  giving a nonzero contribution. Therefore the sum of contributions (B16) and (B17) as  $L \rightarrow \infty$  are given by

$$\begin{aligned} &\lim_{L \rightarrow \infty} \left\{ \sum_{j=2}^{L-1} [B_{j,j+1} \hat{C}_j(\omega) \hat{C}_{j+1}^*(\omega) + B_{j+1,j} \hat{C}_{j+1}(\omega) \hat{C}_j^*(\omega)] \right. \\ &\quad \left. + \sum_{j=2}^L B_{j,j} \hat{C}_j(\omega) \hat{C}_j^*(\omega) \right\} \\ &= \theta \frac{\hat{\beta}^2 \hat{\mu}^2 + \omega^2}{\hat{\mu}^2 + \omega^2} [e^{2\hat{\mu}} - 2e^{\hat{\mu}} \cos \omega + 1]. \end{aligned}$$

Adding together Eqs. (B14)–(B17) we find Eq. (8).

[1] O. N. Bjornstad and B. T. Grenfell, *Science* **293**, 638 (2001).  
 [2] T. Coulson, P. Rohani, and M. Pascual, *Trends Ecol. Evol.* **19**, 359 (2004).  
 [3] P. Rohani, D. J. D. Earn, and B. T. Grenfell, *Science* **286**, 968 (1999).  
 [4] N. C. Grassly, C. Fraser, and G. P. Garnett, *Nature (London)* **433**, 417 (2005).  
 [5] A. J. McKane and T. J. Newman, *Phys. Rev. Lett.* **94**, 218102 (2005).  
 [6] D. Alonso, A. J. McKane, and M. Pascual, *J. R. Soc., Interface* **4**, 575 (2007).  
 [7] H. W. Hethcote and P. van den Driessche, *J. Math. Biol.* **29**, 271 (1991).

[8] W. Wang, *Math. Biosci. Eng.* **3**, 267 (2006).  
 [9] H. W. Hethcote, *Math. Biosci.* **145**, 89 (1997).  
 [10] H. W. Hethcote and P. van den Driessche, *J. Math. Biol.* **40**, 3 (2000).  
 [11] N. M. Ferguson, A. P. Galvani, and R. M. Bush, *Nature (London)* **422**, 428 (2003).  
 [12] T. Gross, Carlos J. Dommar D'Lima, and B. Blasius, *Phys. Rev. Lett.* **96**, 208701 (2006).  
 [13] S. Risau-Gusman and D. H. Zanette, *J. Theor. Biol.* **257**, 52 (2009).  
 [14] L. B. Shaw and I. B. Schwartz, *Phys. Rev. E* **77**, 066101 (2008).  
 [15] O. N. Bjørnstad, B. F. Finkenstädt, and B. T. Grenfell, *Ecol.*

- Monogr. **72**, 169 (2002).
- [16] M. Kuperman and G. Abramson, Phys. Rev. Lett. **86**, 2909 (2001).
- [17] M. Girvan, D. S. Callaway, M. E. J. Newman, and S. H. Strogatz, Phys. Rev. E **65**, 031915 (2002).
- [18] F. S. Vannucchi and S. Boccaletti, Math. Biosci. Eng. **1**, 49 (2004).
- [19] G. Yan, Z. Q. Fu, J. Ren, and W. X. Wang, Phys. Rev. E **75**, 016108 (2007).
- [20] S. Sinha, J. Saramaki, and K. Kaski, Phys. Rev. E **76**, 015101(R) (2007).
- [21] A. Litvak-Hinenzon and L. Stone, J. R. Soc., Interface **6**, 749 (2009).
- [22] R. E. Hope Simpson, Lancet **260**, 549 (1952).
- [23] P. E. Sartwell, Am. J. Hyg. **51**, 310 (1950).
- [24] N. T. J. Bailey, Biometrika **43**, 15 (1956).
- [25] H. W. Hethcote and D. W. Tudor, J. Math. Biol. **9**, 37 (1980).
- [26] Z. Grossman, Theor Popul. Biol. **18**, 204 (1980).
- [27] A. L. Lloyd, Proc. R. Soc. London, Ser. B **268**, 985 (2001).
- [28] A. L. Lloyd, Theor Popul. Biol. **60**, 59 (2001).
- [29] M. Simoes, M. M. Telo da Gama, and A. Nunes, J. R. Soc., Interface **5**, 555 (2008).
- [30] C. R. Cox and H. D. Miller, *The Theory of Stochastic Processes* (Chapman and Hall, London, 1965).
- [31] M. J. Keeling and B. T. Grenfell, Proc. R. Soc. London, Ser. B **269**, 335 (2002).
- [32] H. J. Wearing, P. Rohani, and M. J. Keeling, PLoS Med. **2**, e174 (2005).
- [33] H. T. H. Nguyen and P. Rohani, J. R. Soc., Interface **5**, 403 (2008).
- [34] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
- [35] M. A. Gibson and J. Bruck, J. Phys. Chem. A **104**, 1876 (2000).
- [36] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 1992).
- [37] Wolfram Research, MATHEMATICA, version 6, Wolfram Research Inc., Champaign, Illinois, 2008.
- [38] J. Verdasca, M. M. Telo da Gama, A. Nunes, N. R. Bernardino, J. M. Pacheco, and M. C. Gomes, J. Theor. Biol. **233**, 553 (2005).
- [39] B. Bolker and B. T. Grenfell, Philos. Trans. R. Soc. London, Ser. B **348**, 309 (1995).
- [40] H. Risken, *The Fokker-Planck Equation* (Springer-Verlag, Berlin, 1989).
- [41] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer-Verlag, Berlin, 2004).