

Research Article

Inference for Ecological Dynamical Systems: A Case Study of Two Endemic Diseases

Daniel A. Vasco

Department of Biology, Duke University, Box 90338, Durham, NC 27708, USA

Correspondence should be addressed to Daniel A. Vasco, vasco.daniel.a@gmail.com

Received 2 September 2011; Revised 19 November 2011; Accepted 21 November 2011

Academic Editor: Vikas Rai

Copyright © 2012 Daniel A. Vasco. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A Bayesian Markov chain Monte Carlo method is used to infer parameters for an open stochastic epidemiological model: the Markovian susceptible-infected-recovered (SIR) model, which is suitable for modeling and simulating recurrent epidemics. This allows exploring two major problems of inference appearing in many mechanistic population models. First, trajectories of these processes are often only partly observed. For example, during an epidemic the transmission process is only partly observable: one cannot record infection times. Therefore, one only records cases (infections) as the observations. As a result some means of imputing or reconstructing individuals in the susceptible cases class must be accomplished. Second, the official reporting of observations (cases in epidemiology) is typically done not as they are actually recorded but at some temporal interval over which they have been aggregated. To address these issues, this paper investigates the following problems. Parameter inference for a perfectly sampled open Markovian SIR is first considered. Next inference for an imperfectly observed sample path of the system is studied. Although this second problem has been solved for the case of closed epidemics, it has proven quite difficult for the case of open recurrent epidemics. Lastly, application of the statistical theory is made to measles and pertussis epidemic time series data from 60 UK cities.

1. Introduction

The linking of ecological theory with data is currently a major scientific challenge. Modern methods of data collection and storage are rapidly improving at all levels, from the detailed study of individuals in populations to the distribution of populations and communities over vast landscapes. Despite the ease with which it is possible to develop statistical theory and Bayesian Markov chain Monte Carlo (MCMC) computational statistics for many ecological problems [1], the resolution of many computational issues for these problems remains largely unresolved when fitting dynamical ecological models (either in discrete or continuous time) to large ecological and public health data sets.

In fact, it is possible to discuss many of these computational difficulties using simple stochastic epidemiological models. Epidemiological processes serve as excellent prototypes for exhibiting two major problems of inference that appear in many mechanistic dynamic models. First, the transmission process during an epidemic is only partly

observed. As a result in epidemiology one only records cases and rarely observes the infection time precisely. Second, the official reporting of observations (cases in epidemiology) is typically done not as they are actually recorded, but at some temporal interval over which they have been aggregated. Although these problems have largely been solved for the case of closed epidemics, it has proven quite difficult for the case of open populations that produce recurrent epidemics (endemic diseases) over many generations in continuous time. This is because it is hard to simulate paths that are consistent with the data due to the condition that one must sample from many recorded intervals given the number of infectives in the beginning and at the end of the interval. In general this has proven easier to do for short-duration epidemics because of computational limitations due to data augmentation. As the number of recorded intervals increases the data likelihood computation rapidly becomes intractable or impossible.

In this paper a data augmentation strategy is implemented that allows addressing these problems, is reasonably

straightforward to implement, is fast and accurate for the problem at hand. The basis of the method is a recently proposed Bayesian MCMC algorithm proposed by Wilkinson [2]. This algorithm is used as the computational foundation for inferring parameters using a stochastic epidemiological model: the Markovian susceptible-infected-recovered (SIR) model of epidemiology. The approach used here includes births and deaths as well as immigration of infectives and hence allows modeling of recurrent epidemics and the inference of model parameters for endemic diseases. The computational methods in this paper are largely drawn from recent approaches taken in systems biology for inference of parameters using time series data. In the Results and Discussion (Section 3.4) part of this paper a brief review is made of these computational methodologies. They are compared to the Bayesian approach taken here along with its advantages and limitations.

Most previous work on the SIR using likelihood [3] and Bayesian MCMC [4, 5] has focused on epidemic data sets collected in small closed communities such as households [6, 7] but very little into endemic diseases [8]. Exceptions to this trend are work by Gibson and Renshaw [9] and the more recent work of Cauchemez and Ferguson [8]. The form of the likelihood in the current framework is the same as that presented in O'Neill and Roberts [4] and similar to that of Cauchemez and Ferguson [8] except that in the present study births, deaths, and immigration of infected cases are included in the dynamics. This makes the SIR likelihood used here most similar to that first utilized by Gibson and Renshaw [9]. This assumption is critical in simulating an open population stochastic SIR as an approximate model of endemic diseases. Adding an influx of migrants allows computationally generating patterns of persistent and complex sustained epidemic oscillations [10, 11].

Application of the inference method is made for time series data for two endemic childhood diseases, pertussis and measles. It is shown how to reconstruct stochastic oscillations using simulations and model checking with respect to observed cases. Finally the hypothesis of coherence resonance is investigated and it is shown how it may account for some of the empirically observed patterns of stochastic oscillatory dynamics of the two endemic diseases.

2. Materials and Methods

2.1. SIR Inference: Perfect Information. In this paper a stochastic version of the Kermack-McKendrick susceptible-infectious-recovered (SIR) model [12] will be used to address the inference problem of mechanistic modeling in ecology. As shown below, a structured representation of this model (in fact, any mechanistic model in ecology) can immediately be used to derive a corresponding Markovian stochastic population model. In the deterministic SIR model, there are seven possible events: birth, death (including all possible labeled events for each type of death event), transmission,

recovery, and immigration. The deterministic framework is described by a set of coupled ordinary differential equations:

$$\frac{ds}{dt} = -\frac{\beta}{N}i(t)s(t) + \mu_1N - \mu_2s(t), \quad (1)$$

$$\frac{di}{dt} = -\gamma i(t) + \frac{\beta}{N}i(t)s(t) + \sigma - \mu_3i(t), \quad (2)$$

$$\frac{dr}{dt} = \gamma i(t) - \mu_4r(t) - \sigma. \quad (3)$$

Here, β represents the transmission rate, σ denotes the rate of immigration of infectious individuals, and $1/\gamma$ describes the average infectious period [13]. The immigration term σ is placed in the recovered equation to ensure constant population—a basic assumption underlying the SIR model. The μ_i represent the birth and death rates for each compartment. Note, however, that *per capita* birth and death rates may be assumed to be the same (μ), ensuring a constant long-term population size, N . Also note that here $\mathbf{x} = (x_1(t), x_2(t), x_3(t)) = (s(t), i(t), r(t))$, the 3-dimensional vector of state variables.

Next consider an event-driven model of state change. Define $\alpha = (\beta, \mu_1, \mu_2, \gamma, \mu_3, \sigma, \mu_4)$, which is the 7-dimensional vector of parameters associated with the SI transitions (transitions to the recovered class will be ignored in this paper):

$$\begin{pmatrix} \alpha_1 e_1 \\ \alpha_2 e_2 \\ \alpha_3 e_2 \\ \alpha_4 e_4 \\ \alpha_5 e_5 \\ \alpha_6 e_6 \\ \alpha_7 e_7 \end{pmatrix} = \begin{pmatrix} \text{event 1} \\ \text{event 2} \\ \text{event 3} \\ \text{event 4} \\ \text{event 5} \\ \text{event 6} \\ \text{event 7} \end{pmatrix} = \begin{pmatrix} \text{transmission} \\ \text{birth susceptible} \\ \text{death susceptible} \\ \text{infection} \\ \text{death infected} \\ \text{immigration} \\ \text{death recovered} \end{pmatrix} = \begin{pmatrix} \frac{\beta}{N}si \\ \mu_1N \\ \mu_2s \\ \gamma i \\ \mu_3i \\ \sigma \\ \mu_4r \end{pmatrix}. \quad (4)$$

Define a change in state as if it occurred from some updating rule applied to each possible event. The updating rules are constrained by the structure of the continuous open SIR equations (1)–(4) to specify an association between the event function, $e_j(\mathbf{x})$, and an associated stage-change vector ν_j . Define an *event pathway* vector, $(\mathcal{P}_1, \dots, \mathcal{P}_7)$, where each path \mathcal{P}_j in Table 1 describes a transition event giving rise to integral state changes in S and I . Table 1 shows how events defined by (4) along with the updating of SI states in the SIR model given by (1) and (2) may be used as a template to construct these pathways. The structured representation immediately gives the transition probabilities for the Markovian SIR model [14, 15]. Equations (1)–(3) may now be used to specify the probability event function, $e_j(\mathbf{x})$, and the associated stage-change vector ν_j . For example, since one defines event 1 to represent a transmission event, then $e_1 = P_1(\Delta S^1 = \nu_{11}, \Delta I^1 = \nu_{12})$ has at time t instantaneous rate $e_1(\mathbf{x}, \alpha) = (\beta/N)si$, with $\nu_1 = (\nu_{11}, \nu_{12}) = (-1, 1)$, where P_1 represents the probability of an instantaneous transition for the event path \mathcal{P}_1 . Using the directed network shown in Figure 1, consider event path \mathcal{P}_1 . This is represented in

TABLE 1: Structural representation of SI state-event transitions.

Event path	Parameter	Transition	Flow in node		Flow out node		Flow difference		Change ν_j
	α_j	e_j	S_{in}^j	I_{in}^j	S_{out}^j	I_{out}^j	$\Delta S^j = \nu_{j1}$	$\Delta I^j = \nu_{j2}$	
\mathcal{P}_1	β/N	SI	0	1	1	0	-1	1	ν_1
\mathcal{P}_2	μN	1	1	0	0	0	1	0	ν_2
\mathcal{P}_3	μ	S	0	0	1	0	-1	0	ν_3
\mathcal{P}_4	γ	I	0	0	0	1	0	-1	ν_4
\mathcal{P}_5	μ	I	0	0	0	1	0	-1	ν_5
\mathcal{P}_6	σ	1	0	1	0	0	0	1	ν_6
\mathcal{P}_7	μ	R	0	0	0	0	0	0	ν_7

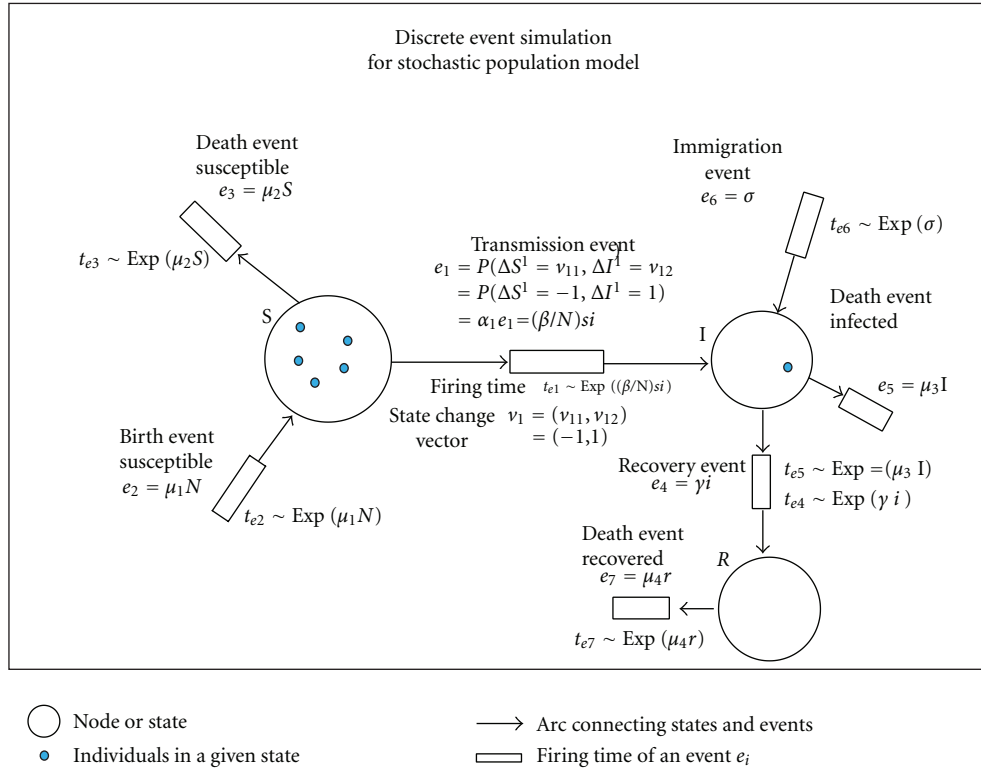


FIGURE 1: Directed network for Markovian SIR dynamics. Each event pathway, \mathcal{P}_j , is represented in the network as the arrow connecting nodes of the random variables, here designated as S , I , and R . The blue dots represent individuals flowing through the network. For example, the flow of individuals out of node with random variable S is represented by the arrow pointing to the box, which gives the firing time to the event and the state-change vector ν_1 associated with the firing of the event.

the network as the arrow connecting nodes of the random S and I variables. The blue dots represent individuals flowing through the network. The flow of individuals out of node with random variable S is represented by the arrow pointing to the box, which gives the firing time to the event and the state-change vector ν_1 associated with the firing of the event. The instantaneous flow (or jumping) between the nodes S and I is determined by $P_1(\Delta S^1 = \nu_{11}, \Delta I^1 = \nu_{12})$. The effect of the firing on an individual in node S is represented by ΔS^1 and corresponds to the first component ν_{11} of the state-change vector ν_1 . The effect on node I is represented by ΔI^1 , which is the second component ν_{12} of state-change vector ν_1 . All of the events in the directed network representing the

open SIR can be treated in the same manner. More generally, one can write

$$e_j(\mathbf{x}, \boldsymbol{\alpha}) dt = \text{probability that } \mathcal{P}_j \text{ event pathway occurs in time } dt. \quad (5)$$

Numerical simulation of stochastic mechanistic models based upon (5) consists of computing the firing of the transitions for each node of the network. The firing of each transition is determined by a random clock running at a time determined by the exponential distribution. For example, in Figure 1 the boxes represent the noisy clocks keeping time

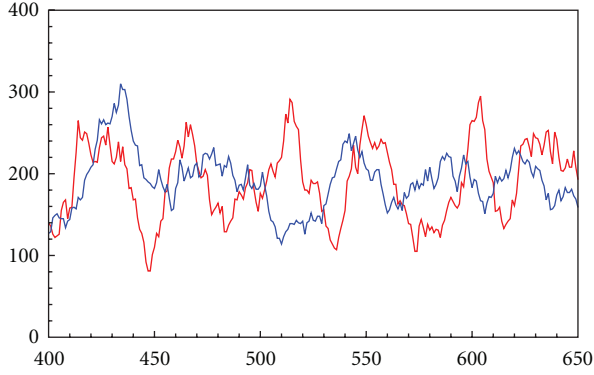


FIGURE 2: A two-hundred-and-fifty-week time series of the number of infected cases from the SIR immigration model discussed in this paper, simulated using the stochastic simulation algorithm (SSA) defined in the supplementary material. y -axis corresponds to observed infected cases. x -axis corresponds to time in weeks. Susceptible cases exhibit a similar complex pattern of noisy oscillations but are not shown. The simulated data are weekly numbers of infected cases; Disease 1 infected cases are shown in blue; Disease 2 infected cases are shown in red. Parameter values used in the simulations: for Disease 1: $\alpha = (\beta, \gamma, \sigma, \mu) = (3.70, .25, .1, .001)$ with $N = 50000$ and for Disease 2: $\alpha = (14.7, .5, .1, .001)$ with $N = 100000$ (see Table 2).

TABLE 2: Baseline disease parameters.

Parameter	Disease 1 Value	Disease 2 Value
N	50000	100000
$1/\mu$	1000 wk	1000 wk
$1/\gamma$	4 wk	2 wk
$1/\sigma$	10 wk	10 wk
β	$3.7^{-\text{wk}} (192^{-\text{yr}})$	$14.7^{-\text{wk}} (764^{-\text{yr}})$

until one goes off in which case a transition is determined by the associated event function.

Figure 2 shows the output for the infected cases of stochastic open SIR. In this section the vector \mathbf{x} represents the sample path for which one has complete information. Assume complete information on the timing and occurrence over a recorded interval of the time series for each individual event propagating through the population. Let $\hat{k} = \sum_{j=1}^{\epsilon} \hat{k}_j$ be the total number of counted events of type \mathcal{P}_j over $[0, T]$. Bookkeep the time and type of event as the set of ordered pairs (t_i, ϵ_i) , where $i = 1, \dots, \hat{k}$, with the t_i in increasing order. Next, consider a recorded event occurring in the ordered interval $[t_{i-1} = t + \tau, t_i = t + \tau + \Delta\tau]$, which was a pathway of type \mathcal{P}_i . In the appendix (Section 1.1) in the supplementary Material available online at doi: 10.1155/2012/390694, it is shown how construction of the likelihood function follows from the stochastic simulation algorithm (SSA) using a factored joint density for any e_{ϵ_i} event tagged with index, ϵ_i , where i is an element of the set consisting of $1, \dots, \epsilon$.

It can also be shown using a factored form of the event function that one can sum over all transitions in the

jump chain resulting from the Kolmogorov forward equation (KFE; see the appendix (Section 1.2)) to obtain,

$$\begin{aligned} \mathcal{L}(\alpha, \mathbf{x}) &= \left\{ \prod_{i=1}^{\hat{k}} \alpha_{\epsilon_i} e_{\epsilon_i}(\mathbf{x}(t_{i-1})) \right\} \times \int_0^T \exp \left\{ -\sum_{i=0}^{\epsilon} \alpha_i e_i dt \right\}, \\ &\propto \prod_{i=1}^{\epsilon} \alpha_i^{\hat{k}_i} \times \exp \left\{ -\alpha_i \int_0^T e_i(\mathbf{x}(t)) dt \right\}, \end{aligned} \quad (6)$$

where ϵ , ϵ_i , \hat{k} , and \hat{k}_i are as defined from above. As shown in the appendix (Sections 1.1–1.3) the standard theory of statistical inference for Markov chains [2, 16, 17] can be applied to simulated Markov processes, to obtain a straightforward, but computationally intensive, maximum likelihood theory for this class of stochastic processes. In fact, these results demonstrate that one can analytically compute closed-form solutions for parameter estimates, since it factors into ϵ independent functions, one for each parameter of an event function and its associated pathway. This gives maximum likelihood estimates of each α_i of the SIR as $\hat{\alpha}_i = k_i / \int_0^T e_i(\mathbf{x}(t)) dt$, for $i = 1, \dots, \epsilon$. This has been demonstrated previously for closed stochastic epidemic models [18, 19]. In this section it has now been shown that similar results hold for *open stochastic endemic disease dynamics*. The factorization will also be utilized in a new way, in a Bayesian context recently advocated by Wilkinson and colleagues [2, 20]. In this case the factorization means that if independent prior distributions are adopted for the parameters this independence will be retained *a posteriori*. Thus, the Bayes theorem may be placed on top of the factorization of likelihood allowing construction of a simulation-based MCMC algorithm for the stochastic SIR. Such an application of this theorem in the SIR case study gives $\alpha_i | \mathbf{x} \sim \Gamma\{a_i + k_i, b_i + \int_0^T e_i(\mathbf{x}(t)) dt\}$, where Γ represents the gamma distribution and $i = 1, \dots, 7$ are indexed by each SIR \mathcal{P}_i specified in Table 1 and Figure 1. However, before this method can be applied to the kind of data obtained from actual epidemics the problem of imperfect observation must be addressed. This will be discussed in the next section of this paper.

2.2. SIR Inference: Imperfect Information

2.2.1. Discrete Data Recording. The previous section dealt with the case of availability of perfect information for an observed sample path. In this section the case of imperfect information, such as when sample paths consist of data obtained on fixed recorded intervals, is considered using the output of the vector \mathbf{x} . Thus, the sampled output vector is now considered to contain only partially observed data. A correction can be computed that depends upon the likelihood of a sample path under the true model and the likelihood of the sample path under an approximate model, which takes into account that data are fixed upon two endpoints. This requires computing the likelihood under an inhomogeneous Poisson process model, which will now be stated (Wilkinson [2], Section 10.2).

For simplicity of notation, it is now assumed that the “true” sample path $\mathbf{x}(t)$ is only observed at times $t = 0$ and $t = 1$. Thus, the data fixed upon two endpoints may now be denoted as $\mathbf{x}(0)$ and $\mathbf{x}(1)$. The complete data likelihood for a discretely sampled trajectory on the interval $[0, 1]$ is then approximately given by

$$\begin{aligned} \mathcal{L}^*(\alpha, \mathbf{x}) &= \left(\prod_{i=1}^k \lambda_{\epsilon_i}(t_i) \right) \exp \left\{ -\frac{1}{2} (e_0(\mathbf{x}(0), \alpha) + e_0(\mathbf{x}(1), \alpha)) \right\}, \end{aligned} \quad (7)$$

where $\lambda_j(t) = (1 - t)e_j(\mathbf{x}(0), \alpha) + te_j(\mathbf{x}(1), \alpha)$, $j = 1, \dots, \epsilon$, and represents the rate of the inhomogeneous Poisson process across the interval.

Using the ratio of likelihoods, $\mathcal{L}/\mathcal{L}^*$, allows one to make a robust statistical decision with respect to accepting or rejecting a discretely sampled time interval.

Using a Poisson approximation allows implementing a very fast stochastic simulation algorithm simply (much faster than the standard SSA) by applying probability functions to deterministic flow rates. This essentially corresponds to computing Euler increments for the τ -leap stochastic simulation method [21]. These computational algorithms are briefly described in the appendix (Sections 1.3–1.5).

MCMC implementation using this framework is reasonably straightforward (see [2], Section 10.3): (a) initialize the algorithm with a valid sample path consistent with the observed data. (b) Sample the SIR parameters from their full conditionals given their current sample paths. (c) For each of the reported time intervals propose new sample paths consistent with the reported endpoints and accept/reject it with the Metropolis-Hastings step. (d) Output MCMC state. Go back to (a). Details of the application of this algorithm to the Markovian SIR are discussed in the appendix (Section 1.6).

2.2.2. Nonobservance of Susceptible Cases. Because numbers of susceptible cases are not available from direct observation they must be reconstructed from the epidemic data. For both the simulation and empirical estimation studies a simple reconstruction method [22] is used. This method utilizes the relationship

$$S_{t+\tau} = S_t - C_{t,\tau} + \bar{C}_{t,\tau}, \quad (8)$$

$$S_0 = 0, \quad (9)$$

where S_t is the number of individuals in the susceptible class, $C_{t,\tau}$ the number of reported cases, and $\bar{C}_{t,\tau}$ the average reported number of cases over the entire data set. Given the case report data the susceptible cases are reconstructed by integrating (8) forward from $t = 0$.

3. Results and Discussion

3.1. Reconstructing Stochastic Oscillations. It has long been a challenge in mathematical epidemiology to understand

the recurrence of epidemic outbreaks and establish an appropriate model that allows studying this phenomenon [23–27]. Recurrent epidemics often exhibit intricate and complex dynamics that cannot easily be studied using deterministic models; demographic stochasticity may play a critical role in determining the outcome of the process especially when the population falls below a certain critical size (the critical community size) [28, 29]. Many recent theoretical studies expanding upon Bartlett’s concept of “intrinsic stochastic oscillations” have assumed that the population persists in a long-term stochastic epidemic state [10, 11, 30]; a similar assumption is made in theoretical studies of complex stochastic oscillations in predator-prey systems [31]. This paper will now explore this scenario and estimate parameters for persistent noisy recurrent epidemics using the data assimilation models described in the previous section.

Parameters were estimated for a time series simulated using the stochastic SIR immigration model described previously in this paper. The parameter vector used for α is shown in Table 2 and is representative of a recurrent childhood disease such as a measles, mumps, or pertussis. Two recurrent epidemic scenarios are explored. These are labeled Disease 1 and Disease 2 in Table 2. City sizes of $N = 50000$ and $N = 100,000$ are assumed along with a life expectancy on the order of 20 years. To model the recurrent nature of such an epidemic, an infective immigration rate of $\sigma = .1$ was assumed, so that there is, on average one new infective arriving every 10 weeks. The numbers of infected cases and susceptible cases are always plotted at weekly intervals in the figures. Likewise the sampling interval used to estimate parameters was always made at weekly intervals. This corresponds to the imperfect observation scenario described in Section 2.2.1. In the next section the scenario in which case reports must be used to reconstruct the susceptible class will be dealt with. An example of a simulated infected cases time series is shown in Figure 2: the blue line is from a simulation using Disease 1 parameter values; the red line is from a simulation using Disease 2 parameter values. Two hundred and fifty weeks of observations of infected cases for the SIR immigration model discussed in this paper are shown. Susceptible cases exhibit a similar complex pattern of noisy oscillations but are not shown in the figure.

Using both infected and susceptible cases time series obtained from the simulations the parameters shown in Table 3 were inferred. Analysis of the MCMC data was accomplished using standard Bayesian data analysis [2, 20, 32–34]. Posterior averages and their standard deviations were used to infer parameters after two million MCMC iterations of the inference algorithm. A burnoff of 100000 iterations was made and iterations were thinned every 100 values.

Figure 4 shows Markov chain traces for five hundred weeks of observations of Disease 2. Rapid convergence of the chain towards a region including the target parameters is seen for all SIR parameters (β, γ, μ) except for the immigration rate σ . The color panel shows how estimation of σ improves as data are added (in the figure a yellow line is used to indicate the results for five hundred weeks of observations, a red line for one thousand weeks of observations, and a

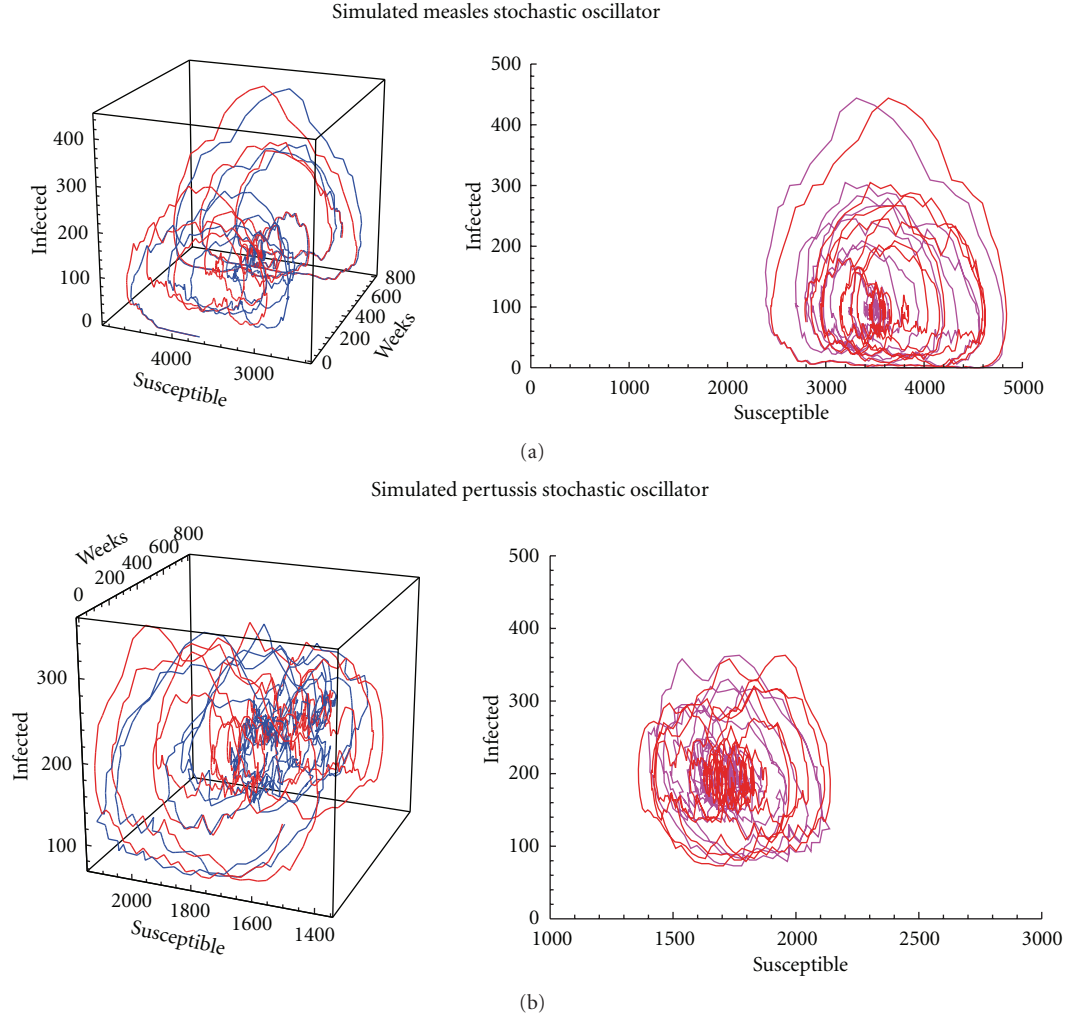


FIGURE 3: Simulation of measles and pertussis as stochastic oscillators with comparison between exact (known) susceptible time series and reconstructed susceptible time series. Parameters used in the simulations are given in Table 2, with pertussis labeled Disease 1 and measles Disease 2. The abbreviation s.o. stands for “stochastic oscillator.” (a) measles stochastic oscillator; (b) pertussis stochastic oscillator.

blue line for ten thousand weeks of observations). Figure 5 shows kernel density estimates for five hundred weeks of observations of Disease 2. The kernel density estimate for migration rate, σ , is for 10000 weeks of observations. Similar results were obtained for Disease 1 type recurrent epidemics (results not shown). All parameters of the epidemic could be estimated for sufficiently long time series (see Table 3).

Finally, it should be pointed out that nearly unbiased estimation of the SIR parameters (β, γ, μ) is sufficient for attractor reconstruction of a persistent recurrent epidemic, at least if the dominant eigenvalue of the point attractor is to be inferred, which is thought to be an important component in driving noisy oscillations of recurrent epidemics in work going back to Soper [23–29].

3.2. Epidemic Inference for 60 UK Cities. In this section parameters are estimated using time series data for 60 UK cities. Pertussis and measles data were obtained using case notification records from the UK Registrar General for England and Wales. Pertussis cases were reported weekly and

biweekly for measles. For both diseases cases reported from the period 1944–1967 were analyzed. City sizes ranged from 10530 (Teignmouth) to 3249440 (London). Reported cases for three UK cities are shown in Figure 7.

Reconstructed susceptible cases (based upon the method described in Section 2.2.2) using simulated measles and pertussis infected time series are shown in Figure 6. Figure 3 shows results from simulation of measles and pertussis stochastic oscillators. Application of this method to perform attractor reconstruction for observed measles time series are shown in Figure 7 for four UK cities. Reasonable similarities were obtained in the comparison between exact (known) susceptible time series versus reconstructed susceptible time series. Parameters used in the simulations are given in Table 2 with pertussis labeled as Disease 1 and measles as Disease 2.

Figure 8 shows the estimates obtained from 60 UK cities for pertussis and measles. Most notable is the large amount of statistical variation seen in the pertussis estimates, particularly in estimates of the duration of infection.

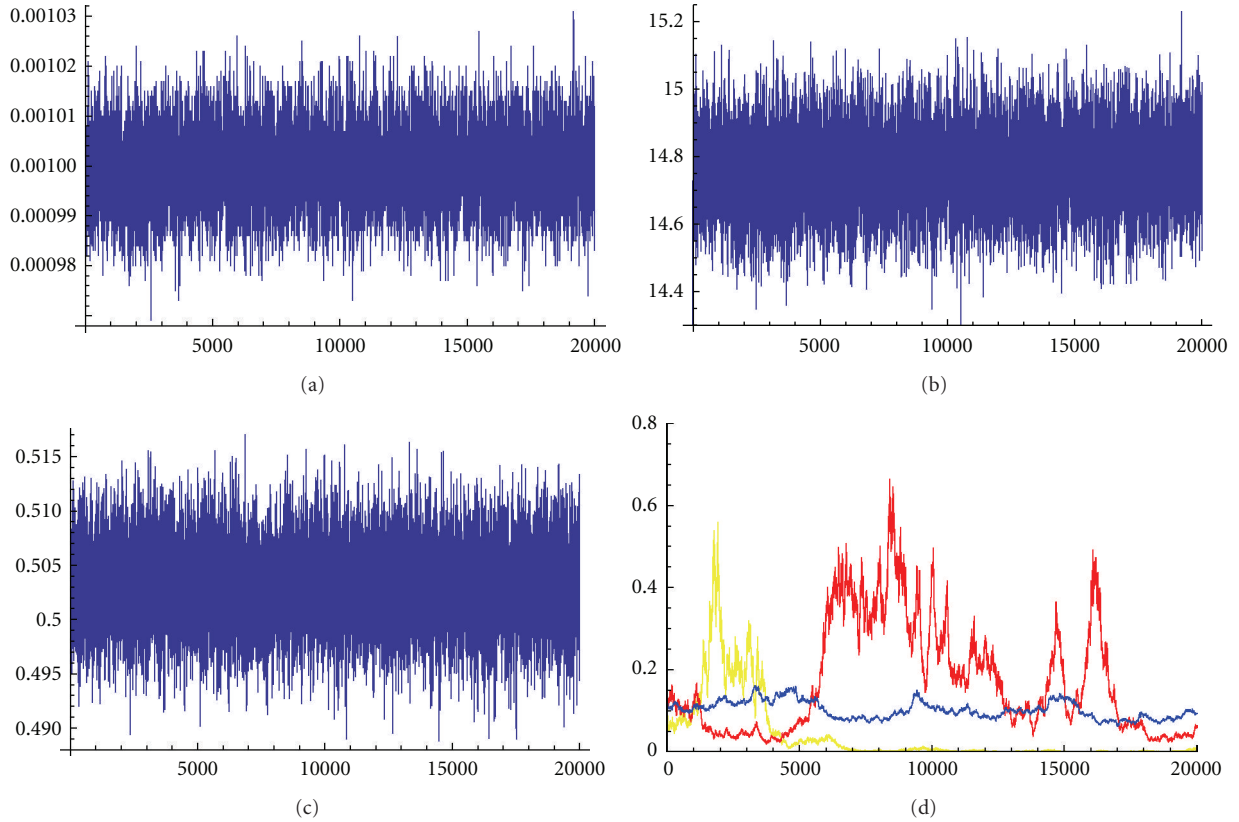


FIGURE 4: Markov chain traces for 500 weeks of observations of Disease 2. The color panel shows how estimation of the migration rate, σ , improves as data are added (yellow line for 500 weeks of observations, red line for 1000 weeks of observations, and blue line for 10,000 weeks of observations). (a) Trace of transmission rate (β); (b) trace of birth rate (N); (c) trace of infection rate (γ); (d) trace of migration rate (σ).

TABLE 3: Posterior estimates of stochastic SIR model.

10000 weeks of observations—Disease 1			
Target	Disease value	Mean	Standard deviation posterior
β	3.70	3.69	.042
γ	.25	.250	.001
σ	.10	.107	.021
μ	.001	.001000	.000004
1000 weeks of observations—Disease 2			
Target	Disease value	Mean	Standard deviation posterior
β	14.7	14.70	.015
γ	.5	.50	.004
σ	.10	.170	.136
μ	.001	.00100	7.18×10^{-6}
10000 weeks of observations—Disease 2			
Target	Disease value	Mean	Standard deviation posterior
β	14.7	14.72	.20
γ	.5	.50	.007
σ	.10	.11	.032
μ	.001	.00100	1.47×10^{-5}

3.3. *Inferring Coherence Resonance.* Coherence resonance occurs when noise is amplified in an otherwise quiescent

system by interaction of the underlying stochasticity of the dynamics with the oscillatory transients of the deterministic dynamics. What has been lacking thus far is a rigorous statistical approach that allows quantifying the theoretical expectations that drive this process using observed time series data. The method developed in this paper is now used to infer endemic sustained oscillations for noisy measles and pertussis epidemics via the mechanism of coherence resonance.

Kuske et al. [11] showed that the Poisson process model of the SIR may be approximated using a stochastic ordinary differential equation with a change of variables. The linearization of scaled model has oscillations that are slowly decaying with unity frequency. Kuske et al. conjecture that solutions of the stochastic SIR model resemble a different approximate model which captures the essence of the full stochastic model. In this stochastic analogue the sustained oscillations have a very particular structure: they are a family of sinusoids modulated by the Ornstein-Uhlenbeck processes. Making this conjecture Kuske et al. [11] derive simple quantitative conditions for the existence of sustained oscillations in noisy time series. Hence, they are able to describe the parameter region for R_0 and γ in detail, including the behavior of the power spectral density of stochastic model and its multiscale approximation. Their parameter space will be taken as the starting point

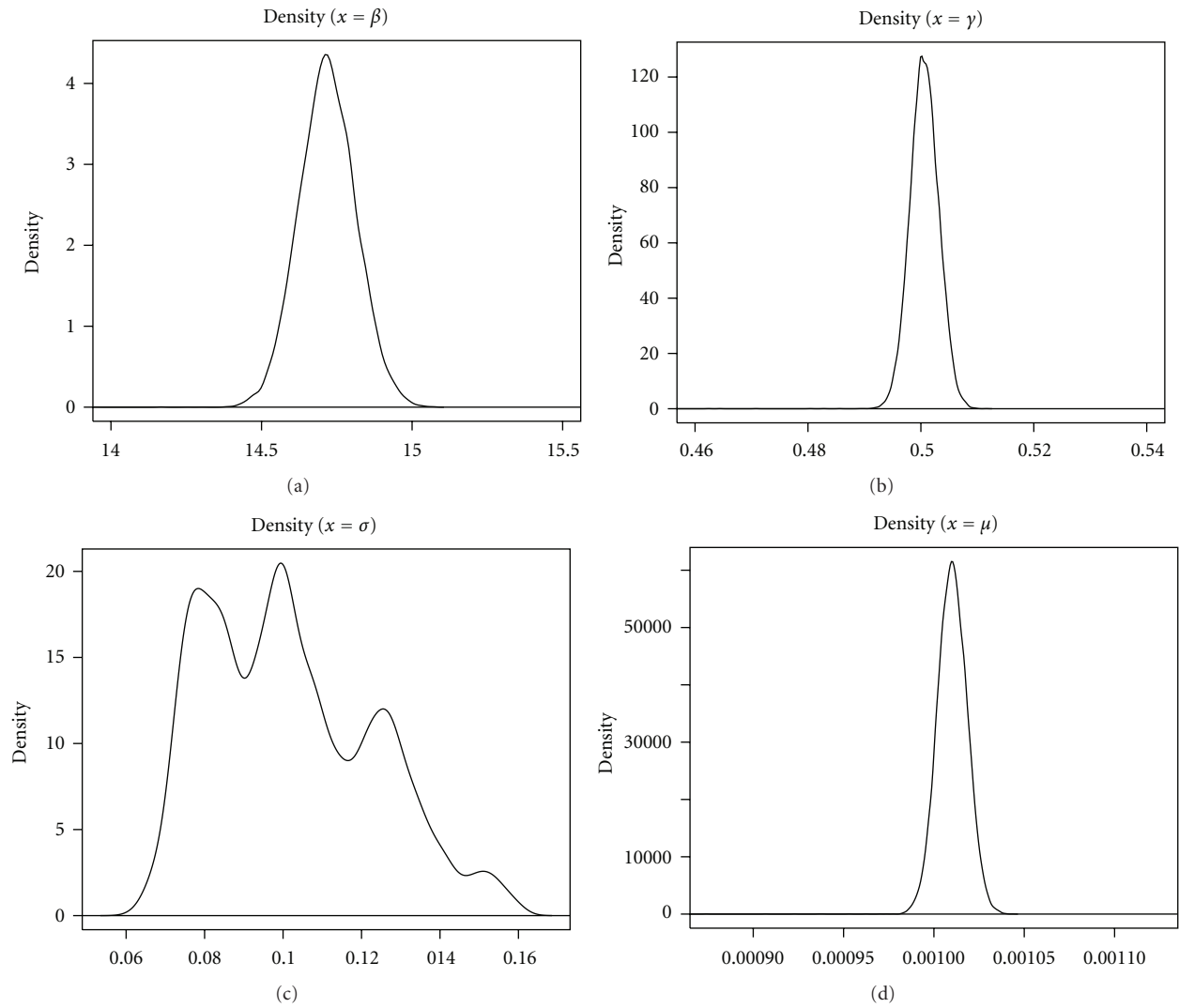


FIGURE 5: Kernel density estimates for 500 weeks of observations of Disease 2. The kernel density estimate for migration rate, σ , is for 10,000 thousand weeks of observations. (b) Density of of transmission rate (β); (a) density of infection rate (γ); (c) density of migration rate (σ); (d) density of birth rate (μ).

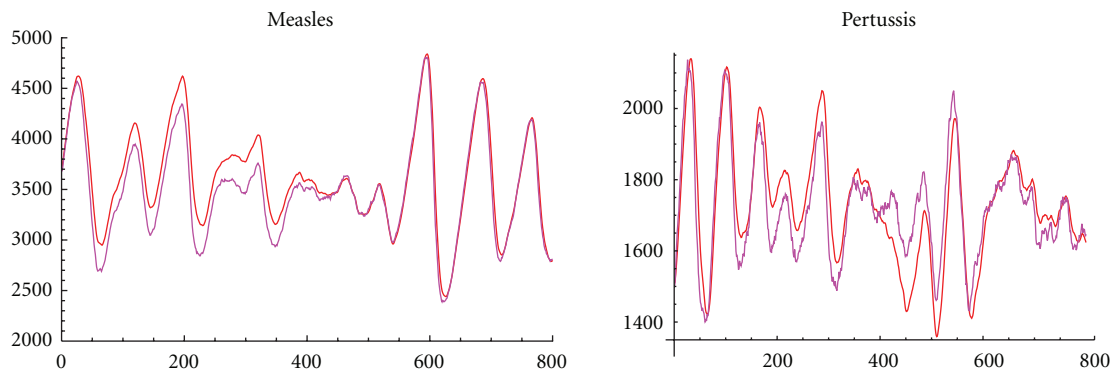


FIGURE 6: Measles and pertussis reconstructed susceptible time series. Cases are on x -axis, weeks are plotted on the x -axis. Blue: susceptible time series. Red: reconstructed susceptible time series.

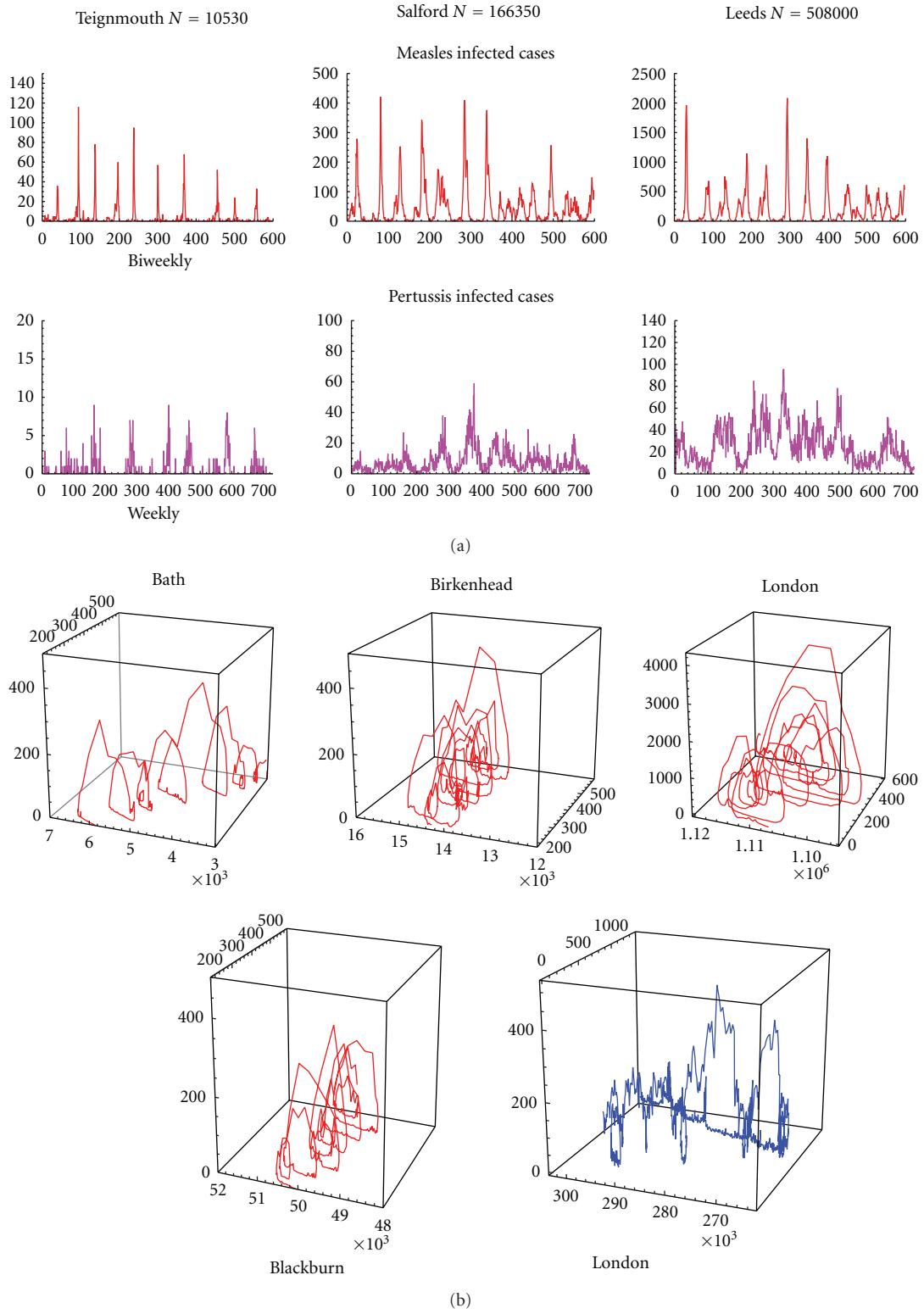


FIGURE 7: (a) Measles and pertussis cases (y -axis) plotted as time series in three UK cities. Measles infected cases were reported as biweekly cases, while pertussis cases were reported weekly. (b) Measles (red) time series as reconstructed stochastic oscillators for four UK cities. The reconstructed oscillator for pertussis is shown in blue for London.

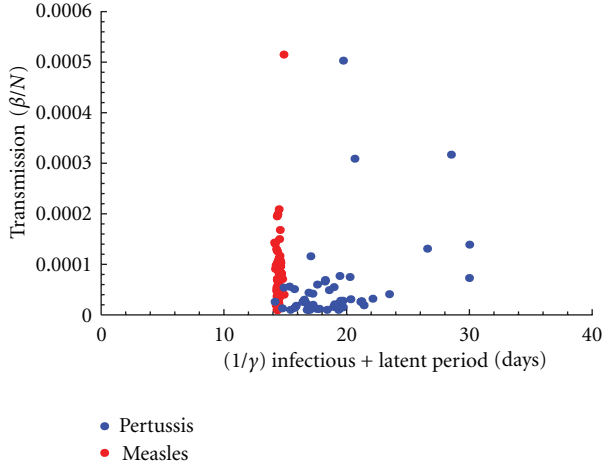


FIGURE 8: Measles and pertussis estimates of transmission (β) and infectious period ($1/\gamma$) epidemic parameters for 60 UK cities.

for formulating the hypothesis of coherence resonance in stochastic epidemics explored in this paper.

Kuske et al. [11] give the biological criterion for sustained oscillations via coherence resonance for the SIR model in terms of two bounds:

$$\epsilon^2 = \frac{R_0}{2} \sqrt{\frac{\mu}{\mu + \gamma} \frac{1}{R_0 - 1}} \ll 1, \quad (10)$$

$$\frac{\delta^2}{2\epsilon^2} = \frac{\mu + \gamma}{4N\mu} \left(1 + \frac{R_0 + 1}{R_0 - 1} + 2 \frac{\mu + \gamma}{\mu(R_0 - 1)} \right) \ll 1.$$

Hence, these bounds can be explored by estimating R_0 and γ that explore this region for the UK measles and pertussis data. Assume that $1/\mu = 70$ is a nuisance parameter. Although σ was estimated, it does not play a role in the following analysis; therefore it is ignored in this section. In addition to μ parameter estimates of interest are those of γ , β , and R_0 which are required to estimate ϵ^2 and $\delta^2/2\epsilon^2$. The major predictions with respect to stochastic amplification in the model of Kuske et al. [11, Page 465] are as follows: (1) for very small values of $\delta^2/2\epsilon^2$ one expects to see very small oscillations. (2) When $\delta^2/2\epsilon^2$ is increased but below one the stochastic fluctuations balance with the deterministic slow decay so that both stochastic and deterministic components interplay to determine the attractor dynamics. (3) When $\delta^2/2\epsilon^2$ is large the stochastic variations govern the dynamics so that an approximation based upon slowly varying modulations is no longer appropriate.

The key results are as follows. Figure 9 shows the estimated variance of the stationary process from measles and pertussis time series. Since this quantity restricts the variance of the stochastic fluctuations relative to the slow time scale it can be used to determine the relative sensitivity of fluctuations on this time scale. It may be observed in Figure 9 that there exist very small estimated values $\delta^2/2\epsilon^2$ for pertussis (blue); hence, one expects to see relatively small very noisy oscillations propagated through the attractor. In this case the demographic noise will not be likely to be amplified optimally with respect to the deterministic

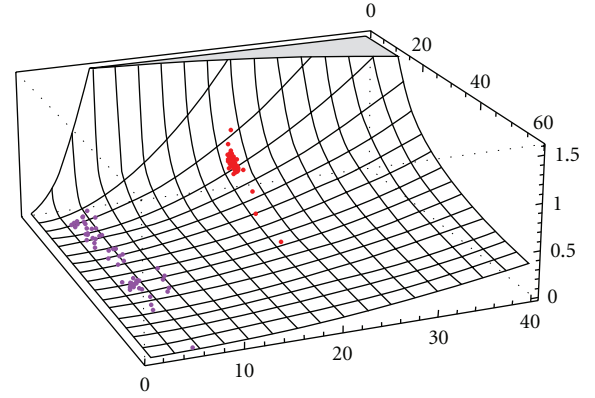


FIGURE 9: Estimated variance of the stationary process from measles and pertussis time series. Plotted on the x -axis is the rate of infection (γ) in years. Plotted on the z -axis is the variance ($\delta^2/2\epsilon^2$). Plotted y -axis is the reproductive rate of the disease (R_0). One observes very small estimated values $\delta^2/2\epsilon^2$ for pertussis (blue), hence one expects to see relatively small very noisy oscillations propagated through the attractor. In contrast, measles (red) exhibits more moderate estimated values $\delta^2/2\epsilon^2$. This implies that the stochastic fluctuations balance with the deterministic slow decay so that both stochastic and deterministic processes contribute the dynamics in terms of producing patterns of coherence resonance.

frequency in the power spectrum and will show more irregular fluctuations due to stochastic amplification of demographic noise. Hence, it may be predicted that the power spectral distribution will not be as sharply peaked and that the multiscale approximation is not as valid for pertussis as for other pathogens. In contrast, Figure 10 shows that for measles (red) one observes more moderate estimated values of $\delta^2/2\epsilon^2$. This implies that the stochastic fluctuations balance with the deterministic slow decay so that both stochastic and deterministic processes contribute the dynamics in terms of producing patterns of coherence resonance. For measles epidemics it is predicted that the power spectral density will have stronger peaks in the vicinity of the deterministic frequency. Measles noisy oscillations are predicted to be better structured and exhibit more coherent cycles around the endogenous period and measles epidemics will exhibit more sensitivity to stochastic amplification. That is they will amplify the noise to generate more regular stochastic cycles in the neighborhood of a fixed frequency.

Figure 10 shows a plot of estimated per year rate of infection (labeled as γ) versus R_0 in analytically predicted bounds expected for multiscale dynamics leading to coherence resonance. In Figure 9 the light green line represents the V-shaped boundary of $\epsilon^2 < .1$ computed using $N = 500000$ and $\mu = 1/55$. This region is approximately the same size when computed for ranges of N between 500000 and 2,000000 [11]. The blue line in Figure 7 represents the contour of the bound $\delta^2/2\epsilon^2 < .2$. Both measles (red) and pertussis (blue) estimates lie well within the bound set by $\delta^2/2\epsilon^2$ for coherence resonance; however, pertussis lies on the boundary of the ϵ^2 bound, which seems to suggest that these epidemics are *not* as likely to exhibit

multiscale dynamics as measles epidemics. There does not appear to exist quite a strong separation between slow and fast time scales in determining pertussis dynamics as there does for measles dynamics. Hence, one expects less coherence and less structured oscillations for pertussis more coherence and structured oscillation for measles epidemics. These results are supported by those observed in Figure 10 and complement each other.

3.4. Systems Biology Approaches to Inference. This paper utilizes a parameter estimation used for mechanistic modeling of biochemical systems [2] to address the important challenge of bridging the gap that exists between mathematical modeling of epidemics and data analysis. In this paper the Bayesian MCMC method has been shown to be useful in bridging such a gap as well as in testing interesting hypotheses regarding the properties of stochastic amplification in epidemics. However, the application of this computational theory in this paper to a simplified open Markovian SIR is really just a first step. But it is an important one and has allowed investigating the properties of data from endemic diseases—a highly nontrivial inference problem in epidemiology. In this section some other, more recent systems biology methods are reviewed and compared to the method used in this paper. Some advantages of systems biology inference methods will be briefly discussed and may be used building upon the results presented in this paper.

The application of computational and mathematical techniques from what has been called algorithmic systems biology [35] to a epidemiological modeling problems will likely prove fruitful. The derivations of stochastic modeling of continuous time processes and the corresponding likelihoods are quite general. However, the approach by Wilkinson [2] and colleagues was a first step in modeling systems in which stochastic effects due to small numbers of molecules or individuals in populations are to be studied. In fact, subsequent studies by the authors focused on inference methods based upon diffusion approximations, which are more tractable and scale up to large systems more easily but are not appropriate for systems in which low densities are common. Applying the Wilkinson Bayesian MCMC Markov jump process approach requires approximating a continuous system using a discrete Poisson approximation. However, as shown in this paper such an approximation does allow obtaining results for endemic diseases which would otherwise be impossible to obtain using earlier algorithms put forward such as that by Gibson and Renshaw [9] for example. Also even as the smaller scale used in this paper it is so computationally intensive it cannot yet be applied to larger scale problems. That is the main reason in this paper that a simplified Markovian SIR model was used. However, even for simplified models one may have the problem of taking long waiting times for rare events. Both simulated and variational maximum likelihood methods in systems biology [36–38] suffer from similar maladies at the Bayesian MCMC methods.

Recent breakthroughs in automated estimation of rare event probabilities in biochemical systems [39–41], however, may allow addressing some of these fundamental problems.

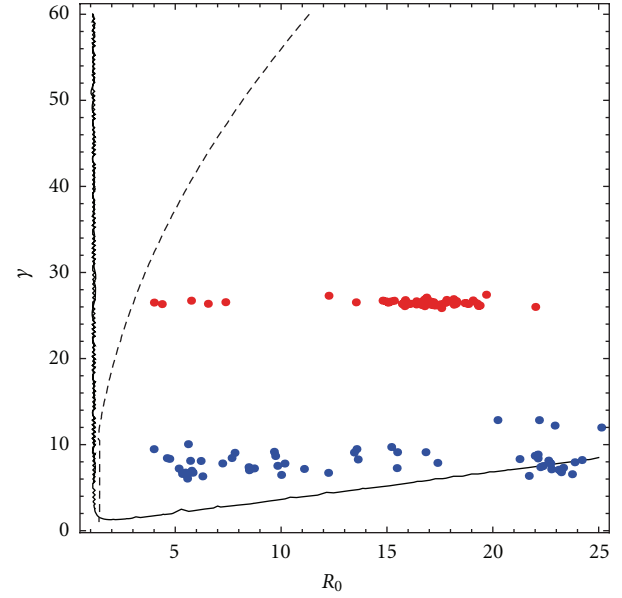


FIGURE 10: Plot of estimated per year rate of infection (labeled gamma on the y -axis) versus R_0 (labeled on the x -axis) in analytically predicted bounds expected for multiscale dynamics leading to coherence resonance. The light green line represents the V-shaped boundary of $\epsilon^2 < .1$ computed using $N = 500000$ and $\mu = 1/55$. This region is approximately the same size when computed for ranges of N between 500000 and 2,000000 [11]. The blue line represents the contour of the bound $\delta^2/2\epsilon^2 < .2$. Both measles (red) and pertussis (blue) estimates lie well within the bound set by $\delta^2/2\epsilon^2$ for coherence resonance; however, pertussis lies on the boundary of the ϵ^2 bound, which seems to suggest that these epidemics are *not* as likely to exhibit multiscale dynamics as measles epidemics. This can be seen visually by comparing the reconstruction of the stochastic oscillators for measles (red) and pertussis (blue), respectively, for London shown in Figure 7.

For the first time an accelerated maximum likelihood estimation for stochastic biochemical systems is in sight that can be based on the continuous time SSA. Construction of inference algorithms based upon these recent studies in systems biology will allow extending the results presented in this paper to more realistic models of the epidemiological process such as including multiple exposed and infected classes. It will also allow including the possibility of disease interactions which, for two diseases can require up to fifty state variables to model [42].

4. Conclusion

In this paper a straightforward Bayesian MCMC methodology for inferring parameters for open SIR models using stochastic simulation is applied to both simulated and observed epidemic time series data. The methods described in this paper are general enough for extension to more complex epidemiological scenarios, which is currently the goal of future work. This is useful because the efficient integration of complex likelihoods for population models is currently an object of intense ongoing research. Analysis

of the data for the methodology developed in this paper is accomplished using standard Bayesian data analysis [2, 20, 32–34].

The results obtained in this paper show how pertussis and measles epidemics behave with respect to the presence of demographic noise. Time series for 60 UK cities were used to estimate epidemiological parameters for these pathogens. A coherence resonance model was fit to the data to infer the role of multiscale effects in producing period and amplitude in the epidemics. It was found that measles appears to fit the model rather well. However, pertussis does not seem to fit the model, and it is predicted that there does not appear to exist quite a strong separation between slow and fast time scales as for pertussis as seems to exist for measles epidemics. Therefore, one expects less coherence and less structured oscillations for pertussis but more coherence and structured oscillation for measles epidemics. The statistical theory developed in this paper was used to investigate coherence resonance of epidemics [10, 11] using empirical time series data. It is hoped that future work will be directed toward extending these results to more complex epidemic modeling [43] such as theory of immune-mediated processes in pathogen interactions [42, 44].

Acknowledgments

This work was partly funded by the University of Missouri and Duke University. The author would like to thank Helen Wearing and Pej Rohani for comments and suggestions on earlier versions of this paper.

References

- [1] J. S. Clark, *Models for Ecological Data*, Princeton University Press, Princeton, NJ, USA, 2007.
- [2] D. J. Wilkinson, *Stochastic Modelling for Systems Biology*, Chapman and Hall, Boca Raton, Fla, USA, 2006.
- [3] H. Andersson and T. Britton, *Stochastic Epidemic Models and Their Analysis*, Springer, Berlin, Germany, 2000.
- [4] P. D. O'Neill and G. O. Roberts, "Bayesian inference for partially observed stochastic epidemics," *Journal of the Royal Statistical Society. Series A*, vol. 162, no. 1, pp. 121–129, 1999.
- [5] P. D. O'Neill, "A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods," *Mathematical Biosciences*, vol. 180, pp. 103–114, 2002.
- [6] N. G. Becker, *Analysis of Infectious Disease Data*, Chapman and Hall, London, UK, 1989.
- [7] N. G. Becker and T. Britton, "Statistical studies of infectious disease incidence," *Journal of the Royal Statistical Society. Series B*, vol. 61, no. 2, pp. 287–307, 1999.
- [8] S. Cauchemez and N. M. Ferguson, "Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in London," *Journal of the Royal Society Interface*, vol. 5, no. 25, pp. 885–897, 2008.
- [9] G. J. Gibson and E. Renshaw, "Estimating parameters in stochastic models using Markov chain methods," *IMA J. Math. Appl. Med. Biol.*, vol. 15, pp. 19–40, 1998.
- [10] D. Alonso, A. J. McKane, and M. Pascual, "Stochastic amplification in epidemics," *Journal of the Royal Society Interface*, vol. 4, no. 14, pp. 575–582, 2007.
- [11] R. Kuske, L. F. Gordillo, and P. Greenwood, "Sustained oscillations via coherence resonance in SIR," *Journal of Theoretical Biology*, vol. 245, no. 3, pp. 459–469, 2007.
- [12] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London Series A*, vol. 115, pp. 700–721, 1927.
- [13] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, Oxford, UK, 1991.
- [14] A. G. McKendrick, "Applications of mathematics to medical problems," *Proceedings of the Edinburgh Mathematical Society*, vol. 44, pp. 98–130, 1926.
- [15] M. S. Barlett, "Some evolutionary stochastic processes," *Journal of the Royal Statistical Society: Series B*, vol. 11, pp. 211–229, 1949.
- [16] P. Billingsley, *Statistical Inference for Markov Processes*, University of Chicago, Chicago, Ill, USA, 1961.
- [17] P. Guttorp, *Stochastic Modeling of Scientific Data*, Chapman and Hall, Boca Raton, Fla, USA, 1995.
- [18] W. N. Rida, "Asymptotic properties of some estimators for the infection rate in the general stochastic epidemic model," *Journal of the Royal Statistical Society: Series B*, vol. 53, pp. 269–283, 1991.
- [19] R. J. Kryscio, "The transition probabilities of the general stochastic epidemic model," *Journal of Applied Probability*, vol. 12, pp. 415–424, 1975.
- [20] R. J. Boys, D. J. Wilkinson, and T. B. L. Kirkwood, "Bayesian inference for a discretely observed stochastic kinetic model," *Statistics and Computing*, vol. 18, no. 2, pp. 125–135, 2008.
- [21] D. T. Gillespie, "Approximate accelerated stochastic simulation of chemically reacting systems," *Journal of Chemical Physics*, vol. 115, no. 4, pp. 1716–1733, 2001.
- [22] G. V. Bobashev, S. P. Ellner, D. W. Nychka, and B. T. Grenfell, "Reconstructing susceptible and recruitment dynamics from measles epidemic data," *Mathematical Population Studies*, vol. 8, no. 1, pp. 1–29, 2000.
- [23] W. H. Hamer, "Epidemic disease in England—the evidence of variability and of persistence of type," *Lancet*, vol. 1, pp. 733–739, 1906.
- [24] H. E. Soper, "The interpretation of periodicity in disease prevalence," *Journal of the Royal Statistical Society: Series B*, vol. 92, pp. 34–73, 1929.
- [25] M. S. Bartlett, *Stochastic Models in Ecology and Epidemiology*, Methuen, London, UK, 1960.
- [26] M. S. Bartlett, "Chance or chaos?" *Journal of the Royal Statistical Society: Series A*, vol. 153, pp. 321–349, 1990.
- [27] N. T. J. Bailey, *The Mathematical Theory of Infectious Disease and Its applications*, Griffin, London, UK, 1975.
- [28] I. Nåsell, "On the time to extinction in recurrent epidemics," *Journal of the Royal Statistical Society. Series B*, vol. 61, no. 2, pp. 309–330, 1999.
- [29] I. Nasell, "Measles outbreaks are not chaotic," in *Mathematical Approaches for Emerging and Reemerging Infectious Disease: Models, Methods and Theory*, Springer, 2002.
- [30] N. G. V. Kampen, *Stochastic Processes in Physics and Chemistry*, North-Holland, Amsterdam, The Netherlands, 1981.
- [31] A. J. McKane and T. J. Newman, "Predator-prey cycles from resonant amplification of demographic stochasticity," *Physical Review Letters*, vol. 94, no. 21, Article ID 218102, pp. 1–4, 2005.
- [32] B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, CRC Press, Boca Raton, Fla, USA, 2000.
- [33] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, CRC Press, Boca Raton, Fla, USA, 2004.

- [34] J. S. Clark, "Why environmental scientists are becoming Bayesians," *Ecology Letters*, vol. 8, no. 1, pp. 2–14, 2005.
- [35] C. Priami, "Algorithmic systems biology," *Communications of the ACM*, vol. 52, no. 5, pp. 80–88, 2009.
- [36] S. K. Poovathingal and R. Gunawan, "Global parameter estimation methods for stochastic biochemical systems," *BMC Bioinformatics*, vol. 11, article 414, 2010.
- [37] T. Tian, S. Xu, J. Gao, and K. Burrage, "Simulated maximum likelihood method for estimating kinetic rates in gene expression," *Bioinformatics*, vol. 23, no. 1, pp. 84–91, 2007.
- [38] Y. Wang, S. Christley, E. Mjolsness, and X. Xie, "Parameter inference for discretely observed stochastic kinetic models using stochastic gradient descent," *BMC Systems Biology*, vol. 4, article 99, 2010.
- [39] B. J. Daigle, M. K. Roh, D. T. Gillespie, and L. R. Petzold, "Automated estimation of rare event probabilities in biochemical systems," *Journal of Chemical Physics*, vol. 134, no. 4, Article ID 044110, 2011.
- [40] H. Kuwahara and I. Mura, "An efficient and exact stochastic simulation method to analyze rare events in biochemical systems," *Journal of Chemical Physics*, vol. 129, no. 16, Article ID 165101, 2008.
- [41] D. T. Gillespie, M. Roh, and L. R. Petzold, "Refining the weighted stochastic simulation algorithm," *Journal of Chemical Physics*, vol. 130, no. 17, Article ID 174103, 2009.
- [42] D. A. Vasco, H. J. Wearing, and P. Rohani, "Tracking the dynamics of pathogen interactions: modeling ecological and immune-mediated processes in a two-pathogen single-host system," *Journal of Theoretical Biology*, vol. 245, no. 1, pp. 9–25, 2007.
- [43] M. J. Keeling and P. Rohani, *Modelling Infectious Diseases*, Princeton University Press, Princeton, NJ, USA, 2007.
- [44] P. Rohani, C. J. Green, N. B. Mantilla-Beniers, and B. T. Grenfell, "Ecological interference between fatal diseases," *Nature*, vol. 422, no. 6934, pp. 885–888, 2003.