National Institute for Applied Statistics Research Australia

The University of Wollongong

Working Paper

09-14

Bayesian Hierarchical Statistical SIRS Models

Lili Zhuang, Noel Cressie

Bayesian Hierarchical Statistical SIRS Models

Lili Zhuang · Noel Cressie

Abstract The classic SIR (susceptible-infectious-recovered) model, has been used extensively to study the dynamical evolution of an infectious disease in a large population. The SIR-susceptible (SIRS) model is an extension of the SIR model to allow modeling imperfect immunity (those who have recovered might become susceptible again). SIR(S) models assume observed counts are "mass balanced." Here, mass balance means that total count equals the sum of counts of the individual components of the model. However, since the observed counts have errors, we propose a model that assigns the mass balance to the hidden process of a (Bayesian) hierarchical SIRS (HSIRS) model. Another challenge is to capture the stochastic or random nature of an epidemic process in a SIRS. The HSIRS model accomplishes this through modeling the dynamical evolution on a transformed scale. Through simulation, we compare the HSIRS model to the classic SIRS (CSIRS) model, a model where it is assumed that the observed counts are mass balanced and the dynamical evolution is deterministic.

Keywords: mass balance, disease dynamics, epidemic model, influenza, HSIRS

Lili Zhuang
Department of Statistics, The Ohio State University, Columbus, OH 43210
Tel.: +614-29-22866
E-mail: zhuang.11@buckeyemail.osu.edu

Noel Cressie
NIASRA, School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, NSW 2522, Australia

1 Introduction

A pandemic (e.g., caused by influenza viruses such as H1N1, H5N1) is an epidemic of an infectious disease spreading through human populations across a large region (e.g., Potter, 2001). Recently, the risk of pandemic influenza has been a significant public-health concern, and much attention has been paid to achieve more precise and timely estimates and predictions of influenza activity.

Compartment epidemic models, such as the SIR (susceptible-infectious-recovered) model, have been widely used to study the dynamical evolution of an infectious disease in a large population. The SIR model was first proposed by Kermack and McKendrick (1927) to explain plague and cholera epidemics, and it was extended to the SIR-susceptible (SIRS) model to allow imperfect immunity (Kermack and McKendrick, 1932; Kermack and McKendrick, 1933; Dushoff et al., 2004). Assume that at any given time t, a fixed population can be split into three compartments: $S(t)$ denotes the number of susceptibles; $I(t)$ denotes the number of infectious; and $R(t)$ denotes the number of "recovereds" (which includes deaths). Then in an SIRS model, the dynamical process is captured through the following set of deterministic (i.e., non-stochastic) nonlinear ordinary differential equations (ODEs):

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta SI + \phi R, \tag{1}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta SI - \gamma I, \tag{2}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I - \phi R, \tag{3}$$

In (1)-(3), β denotes the transmission rate (also referred as the contact or infection rate) per unit time, which can be expressed as the fraction of contacts between susceptible individuals and infectious individuals that result in an infection. Further, γ denotes the rate of "recovery" per unit time, which is the rate at which infectious individuals are removed from being infectious due to recovery (or death); then $\frac{1}{\gamma}$ is the average duration of the infectious period. Finally, ϕ denotes the rate of loss of immunity of recovered individuals per unit time, which is the rate at which recovered individuals become susceptible again (Anderson and May, 1991; Hethcote, 2000); then $\frac{1}{\phi}$ is the average duration of the immunity period.

Notice that by adding equations (1)-(3), we can obtain

$$\frac{\mathrm{d}S}{\mathrm{d}t} + \frac{\mathrm{d}I}{\mathrm{d}t} + \frac{\mathrm{d}R}{\mathrm{d}t} = 0. \tag{4}$$

Thus, the model postulates a fixed total population, N, without entry and exits of demographic type (i.e, there are no births or deaths from causes other than the disease itself). Clearly, it represents a short-term phenomenon where

the total number of people in all three compartments together is constant. Thus, (4) implies that for t in an interval of $\mathbb{R}_+$,

$$S(t) + I(t) + R(t) = N. \tag{5}$$

This assumption (5) is commonly referred to as *mass balance* (Reluga, 2004). As discussed above, the SIRS model in (1)-(3) assumes a fraction ϕ of members of the recovered class can rejoin the susceptible class; the traditional SIR model is obtained when $\phi = 0$.

A discretized version of (1)-(3) can be written as a set of deterministic difference equations. That is, when time t is discrete (in units of Δ days),

$$S(t+1) = S(t) - \beta S(t)I(t)\Delta + \phi R(t)\Delta, \tag{6}$$

$$I(t+1) = I(t) + \beta S(t)I(t)\Delta - \gamma I(t)\Delta, \tag{7}$$

$$R(t+1) = R(t) + \gamma I(t)\Delta - \phi R(t)\Delta; \tag{8}$$

and the discrete-time mass-balance constraint is,

$$S(t) + I(t) + R(t) = N, \tag{9}$$

where here $t = 1, 2, ..., T$.

In past decades, the deterministic SIR model and its various extensions (e.g., SIRS, SIR including birth and death rates, migration, etc.), have been used extensively for infectious-disease estimation and prediction in large and well mixed populations (e.g., Schenzle, 1984; Anderson and May, 1991; Keeling and Rohani, 2007). The *classic SIRS (CSIRS) model* given by (1)-(3) or (6)-(8) is appealing because of its straightforward modeling strategy and its easily interpretable parameters. However, there are various sources of uncertainty in the model: First, there is uncertainty in the counts $\{S(t), I(t), R(t)\}$ themselves; that is, the counts in the compartments are observed with error. Second, the rather simple model (1)-(3) (or (6)-(8) in the discrete-time setting) may not capture the uncertainties in the hidden dynamical epidemic process, such as the uncertainties caused by the presence of heterogeneous populations; and third, the values of the parameters β, γ, and ϕ are uncertain.

A variety of stochastic models have been developed recently, through a probabilistic mechanism that involves a Markov chain of SIR states (e.g., Bailey, 1975; Andersson and Britton, 2000; Allen, 2003; Xu et al., 2007). Some more recent stochastic models involve complex networks (e.g., Halloran et al., 2002; Zhou et al., 2006; Volz, 2008) or drug resistance (e.g., Chao et al., 2012) to avoid the assumption of homogenous mixing. However, these stochastic models ignore the noisy nature of data, and they apply mass balance to the *observed* counts rather than the true counts. Furthermore, these models typically rely on many carefully chosen parameters, such as transmission rates, recovery rates, and so forth in heterogeneous populations; uncertainty in where the model's parameter vector is located in the parameter space is not accounted for. Our

strategy is to deal with each source of uncertainty using a Bayesian hierarchical statistical model, where the counts are observed with error and the dynamical evolution embodied in (6)-(8) is stochastic.

Bayesian percolation models have proven popular for modeling spatio-temporal dynamical processes (e.g., Catterall et al., 2012; Gibson et al., 2006) and have been applied to epidemics (e.g., Cook et al., 2007), but they ignore the true process hidden behind the noisy data. More recent Bayesian hierarchical models, which are widely used for mapping non-infectious diseases, aim to capture the true spatial process (e.g., Besag et al., 1991; Carlin and Banerjee, 2002), but their process models and parameter models are not appropriate for epidemics. Those that do have a dynamical spatial statistical component have not generally been parameterized in terms of the interpretable components of the epidemic (e.g., Mugglin et al., 2002; Wood, 2010).

Recently, partially observed nonlinear stochastic dynamical systems (also know as partially observed Markov processes, or state-space models) have been used extensively for infectious-disease estimation and prediction. A wide range of inference techniques have been proposed and implemented in the R statistical language as part of the package pomp (http://cran.at.r-project.org/web/packages/pomp/), such as nonlinear forecasting (e.g., Kendall et al., 1999; Kendall et al., 2005), iterated filtering (Ionides et al., 2006; King et al., 2008; He et al., 2010), and approximate Bayesian particle filtering (e.g., Liu and West, 2001; Arulampalam et al., 2001; Dukić et al., 2012; Rasmussen et al., 2011). Some more recent state-space models involve spatial components in the dynamical process model (e.g., Patterson et al., 2008). Some of these models are not appropriate for modeling epidemic flows (e.g., Kendall et al., 1999; Kendall et al., 2005; Patterson et al., 2008). Those that are extensions of classic compartment epidemic models (e.g., the SIR model and the SEIR model) and that do pay attention to the underlying true process hidden behind the noisy data, either ignore the source of variation that captures randomness in the (hidden) epidemic process (e.g., Rasmussen et al., 2011), or they do not preserve the mass-balance property (e.g., Liu and West, 2001; Dukić et al., 2012), which may introduce biased results. Recent extensions to stochastic models with a master equation have similar problems with mass balance (e.g., Alonso et al., 2007).

Notice that imposing the mass balance appropriately to improve accuracy of estimating counts in all compartments is important. If one only pays attention to estimating the infectious population (I) and ignores modeling the susceptible population (S) and the recovered population (R) (for example, simply using a value of R from the literature and letting $S = N - I - R$, as if it were observed), it may result in biased estimates of important model parameters (e.g., the recovery rate γ) and finally lead to bias in the estimation of I. Further, bias in γ may lead to bias in the reproduction number, R_0 (defined in Section 4.1), which in turn may affect the estimate of the proportion of susceptibles that need to be vaccinated to achieve "herd immunity" (Fine et al.,

2011). Accurate estimates to support public-health decisions are critical for disease control and prevention.

In this article, we return to the classic SIRS (CSIRS) model (6)-(8) for motivation, and we propose a discrete-time, mass-balanced (Bayesian) hierarchical SIRS, or HSIRS, model, which is based directly on counts and imposes mass balance appropriately on the underlying true counts, rather than on the observed counts. Our model captures the randomness in the epidemic process by assuming that the dynamical process occurs on a log-odds-ratio scale, transformed from the scale where the true counts are mass-balanced. This new dynamical approach to infectious-disease modeling also shows how infectious diseases in *heterogeneous* populations could be modeled hierarchically.

In Section 2, we propose the HSIRS model for infectious-disease data. The actual computations associated with the posterior analysis involve local linearization of difference equations; see Section 3. In Section 4, we simulate in discrete time, a dataset from the HSIRS model and from a CSIRS model that is modified to incorporate observation error, and then we infer all unknowns of the models through Markov chain Monte Carlo (MCMC) analysis. Comparisons are given to CSIRS-model-based inference. In Section 5, we extend the example to a carefully designed simulation experiment with sufficient replication to make definitive comparisons between the HSIRS model and the CSIRS model. Discussion and conclusions are given in Section 6. Some technical derivations are given in the Appendix.

2 Bayesian Hierarchical Statistical SIRS (HSIRS) Model

We assume that there is a true (unobserved) process underlying the observed epidemic counts, which we incorporate into the framework of a Bayesian hierarchical statistical model. This typically consists of three components: the data model (i.e., the conditional distribution of the data given hidden processes and parameters); the process model (i.e., the conditional distribution of the hidden processes given parameters); and the parameter model (i.e., the prior distribution of the parameters).

2.1 Data Model

We model the raw counts directly rather than modeling the rates derived from the counts (e.g., Dukić et al., 2012, use Gaussian distributions to model the rates) and assume that the data model consists of (conditionally) independent Poisson distributions evolving at discrete time points. That is, for time points $t = 1, 2, ..., T$, in units of Δ days, the data model is

$$Z_S(t)|P_S(t) \sim \text{ind. Poisson}(\lambda_N P_S(t)), \tag{10}$$

independent from

$$Z_I(t)|P_I(t) \sim \text{ind. Poisson}(\lambda_N P_I(t)), \tag{11}$$

where $Z_S(t)$ and $Z_I(t)$ are the observed number of susceptible and infectious individuals at time t, respectively; "ind." is shorthand for "independent"; λ_N denotes the true total population count, and $P_S(t)$ and $P_I(t)$ are the underlying true rates of susceptible and infectious individuals at time t, respectively. Since λ_N is known from demography and $\lambda_N = Z_S(t) + Z_I(t) + Z_R(t)$, then $Z_R(t)$ follows. Thus, the data are $\{(Z_S(t), Z_I(t)) : t = 1, 2, ..., T\}$.

2.2 Process Model

Recall the discrete-time CSIRS model defined by (6)-(8), which assumes that mass balance happens on the *observed* population. However, the appropriate place to model mass balance is on the true (hidden) process. That is, for $t = 1, 2, ...T$, we have

$$\lambda_S(t) + \lambda_I(t) + \lambda_R(t) = \lambda_N, \tag{12}$$

where $\lambda_S(t)$, $\lambda_I(t)$, and $\lambda_R(t)$ are the underlying true (but hidden) counts of susceptible, infectious, and recovered individuals at time t, respectively. Now define the true (hidden) rates, $P_S(t)$, $P_I(t)$, and $P_R(t)$, via

$$\lambda_S(t) \equiv \lambda_N P_S(t), \tag{13}$$

$$\lambda_I(t) \equiv \lambda_N P_I(t), \tag{14}$$

$$\lambda_R(t) \equiv \lambda_N P_R(t), \tag{15}$$

where $P_R(t)$ denotes the underlying true rate of recovered individuals at time t. Then by substituting (13)-(15) into (12), it is straightforward to see that the mass balance in (12) can be rewritten as,

$$P_S(t) + P_I(t) + P_R(t) = 1, \tag{16}$$

and hence for $t = 1, 2, ...T$,

$$P_R(t) = 1 - P_S(t) - P_I(t). \tag{17}$$

Recall that the difference equations defined in (6)-(8) give easily interpretable dynamics, in which individuals move from the susceptible state, to the infectious state, then to the recovered state (and some individuals may become susceptible again). We wish to model this "SIRS flow" on the hidden process where t is discrete (in units of Δ days). We do this by deriving a set of deterministic difference equations on $\lambda_S(t)$, $\lambda_I(t)$, and $\lambda_R(t)$. That is, for $t = 1, 2, ...,$ our process model is,

$$\lambda_S(t+1) = \lambda_S(t) - \beta \Delta \lambda_S(t) \lambda_I(t) + \phi \Delta \lambda_R(t), \tag{18}$$

$$\lambda_I(t+1) = \lambda_I(t) + \beta\Delta\lambda_S(t)\lambda_I(t) - \gamma\Delta\lambda_I(t), \tag{19}$$

$$\lambda_R(t+1) = \lambda_R(t) + \gamma\Delta\lambda_I(t) - \phi\Delta\lambda_R(t). \tag{20}$$

In (18)-(20), the SIRS flow has been preserved, and the rate parameters β, ϕ, and γ are in units of per day (d^{-1}).

According to the definition of $\lambda_S(t)$, $\lambda_I(t)$, and $\lambda_R(t)$ in (13)-(15), equations (18)-(20) can be rewritten in terms of the true proportions, $P_S(t)$, $P_I(t)$, and $P_R(t)$:

$$P_S(t+1) = P_S(t) - \beta\lambda_N P_S(t)P_I(t)\Delta + \phi P_R(t)\Delta, \tag{21}$$

$$P_I(t+1) = P_I(t) + \beta\lambda_N P_S(t)P_I(t)\Delta - \gamma P_I(t)\Delta, \tag{22}$$

$$P_R(t+1) = P_R(t) + \gamma P_I(t)\Delta - \phi P_R(t)\Delta. \tag{23}$$

Deterministic equations, such as those in (21)-(23), are unable to capture the uncertainties in the hidden epidemic model. Possible sources of uncertainty are the presence of heterogeneous populations and the existence of other categories such as *exposed* individuals. To handle the complexity while still preserving the mass balance, we apply the logit transformation to the true rates, which changes the scale of variability from $[0, 1]$ to $(-\infty, \infty)$. That is, for $t = 1, 2, ...,$ define

$$W_S(t) \equiv \log\left(\frac{P_S(t)}{P_R(t)}\right), \tag{24}$$

$$W_I(t) \equiv \log\left(\frac{P_I(t)}{P_R(t)}\right), \tag{25}$$

where $W_S(t)$ and $W_I(t)$ are the log odds ratios of susceptible-over-recovered populations, and infectious-over-recovered populations, respectively, at time t. On the odds-ratio scale ($\mathbf{W}$-scale), we construct the process model in terms of $\mathbf{W}(t) \equiv (W_S(t), W_I(t))'$:

$$\mathbf{W}(t+1) = \boldsymbol{\mu}^W(t) + \boldsymbol{\xi}(t+1), \tag{26}$$

for discrete time $t = 1, 2, ...,$ in units of Δ days. We now discuss each of the components of (26), in turn. The vector $\boldsymbol{\mu}^W(t) \equiv (\mu_S^W(t), \mu_I^W(t))'$ is the dynamical process that captures the temporal dependence. In Appendix A.1, we derive the nonlinear dynamical structure of $\boldsymbol{\mu}^W(t)$ using (21)-(25). This derivation retains the SIRS flow on the hidden process; that is, for discrete time $t = 1, 2, ...,$ in units of Δ days,

$$
\begin{aligned}
\mu_S^W(t) = \ & W_S(t) \\
& + \log\left[1 + \frac{\phi\Delta}{\exp\left(W_S(t)\right)} - \frac{\beta\lambda_N\exp\left(W_I(t)\right)\Delta}{1 + \exp\left(W_S(t)\right) + \exp\left(W_I(t)\right)}\right] \\
& + \log\left[\frac{1}{1 + \gamma\exp\left(W_I(t)\right)\Delta - \phi\Delta}\right],
\end{aligned}
\tag{27}
$$

$$
\begin{aligned}
\mu_I^W(t) = {} & W_I(t) \\
& + \log\left[1 - \gamma\Delta + \frac{\beta\lambda_N\exp\left(W_S(t)\right)\Delta}{1 + \exp\left(W_S(t)\right) + \exp\left(W_I(t)\right)}\right] \\
& + \log\left[\frac{1}{1 + \gamma\exp\left(W_I(t)\right)\Delta - \phi\Delta}\right],
\end{aligned}
\tag{28}
$$

where recall that β, γ, and ϕ, are the transmission rate, recovery rate, and loss-of-immunity rate per day, respectively.

We denote the vector $\boldsymbol{\xi}(t) \equiv (\xi_S(t), \xi_I(t))'$, to be the small-scale variation that captures the uncertainties in the hidden epidemic process. For $t = 1, 2, ...$, we define

$$
\boldsymbol{\xi}(t) \sim \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}}(t)),
\tag{29}
$$

a multivariate normal (MVN) distribution with mean $\mathbf{0}$ and diagonal covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}(t) \equiv \mathrm{diag}(\sigma_{\xi_S}^2(t), \sigma_{\xi_I}^2(t))$, with nonnegative variance components, $\sigma_{\xi_S}^2(t)$ and $\sigma_{\xi_I}^2(t)$. For the sake of simplicity, in this article we assume that $\sigma_{\xi_S}^2(t) = \sigma_{\xi_S}^2$ and $\sigma_{\xi_I}^2(t) = \sigma_{\xi_I}^2$, for all $t = 1, 2, ...$ Notice that the (negative) covariation of $W_S(t)$ and $W_I(t)$ (defined in (24) and (25)) due to the mass-balance constraint in (16) has been structured through the covariance of the nonlinear dynamical process $\boldsymbol{\mu}^W(t)$. Therefore, it is reasonable to assume that $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}(t)$, which is the covariance of the *conditional* distribution, $[\mathbf{W}(t+1)|\boldsymbol{\mu}^W(t)]$, has a diagonal structure. Also notice that we can still preserve mass balance while adding the small-scale variation, $\boldsymbol{\xi}(t)$, on the $\mathbf{W}$-scale, because the mass balance is preserved when

$$
P_S(t) = \frac{\exp(W_S(t))}{1 + \exp(W_S(t)) + \exp(W_I(t))},
$$

$$
P_I(t) = \frac{\exp(W_I(t))}{1 + \exp(W_S(t)) + \exp(W_I(t))},
$$

$$
P_R(t) = \frac{1}{1 + \exp(W_S(t)) + \exp(W_I(t))}.
$$

Finally, notice that these equations do not put a constraint on the range of the values $W_S(t)$ and $W_I(t)$, so adding the error vector $\boldsymbol{\xi}(\cdot)$ in (26) will not destroy the balance. However, if an error vector is added directly on the $\mathbf{P}$-scale, the mass balance constraint (16) is difficult to retain.

The strategy of transforming from the hidden proportion scale ($\mathbf{P}$-scale) to the hidden log-odds-ratio scale ($\mathbf{W}$-scale) and adding the small-scale variation on the $\mathbf{W}$-scale rather than on the $\mathbf{P}$-scale, is key to retaining the mass-balance constraint while allowing flexible SIRS flow to be handled. To our knowledge, this is a new approach to infectious-disease modeling; other Bayesian approaches (e.g., Dukić et al., 2012) do not preserve mass balance after building uncertainties into the process model.

2.3 Parameter Model

To complete the model, we now specify the joint prior distribution for the parameters, which includes the transmission rate per unit time, β; the rate of recovery per unit time, γ; the loss-of-immunity rate per unit time, ϕ; and variance components, $\{\sigma_{\xi_S}^2\}$ and $\{\sigma_{\xi_I}^2\}$. Notice that the difference equations in (6)-(8) impose a natural constraint on β; that is, for any time $t = 1, 2, ...,$

$$\beta S(t)I(t) \leq N - (1 - \phi)R(t), \tag{30}$$

because the number of individuals that become infectious at a certain time t should be less than or equal to the total number that could be infected at that time. Hence,

$$0 \leq \beta \leq \frac{N - (1 - \phi)R(t)}{S(t)I(t)}. \tag{31}$$

In the context of the HSIRS model, this amounts to ensuring that β is bounded above by a hyperparameter, $\beta_{\max}$. Furthermore, it is straightforward to see that $\gamma \in [0, 1]$ and $\phi \in [0, 1]$, due to their definition. Typically, information about the recovery rate γ is easier to obtain than the other rate parameters; we apply a logit transformation to γ,

$$\theta_\gamma \equiv \log\left(\frac{\gamma}{1 - \gamma}\right), \tag{32}$$

and assign a Gaussian prior to θ_γ.

Assuming statistical independence of parameters and using $[Y]$ as generic notation for the probability distribution of Y, we assume that the parameter model can be written as

$$[\beta, \theta_\gamma, \phi, \sigma_{\xi_S}^2, \sigma_{\xi_I}^2] = [\beta][\theta_\gamma][\phi][\sigma_{\xi_S}^2][\sigma_{\xi_I}^2], \tag{33}$$

where we specify the prior distributions of individual parameters as follows:

$$\beta \sim \text{Uniform}[0, \beta_{max}],$$

$$\theta_\gamma \sim \text{Normal}(\mu_{\theta_\gamma}, \sigma_{\theta_\gamma}^2),$$

$$\phi \sim \text{Uniform}[0, 1],$$

$$\sigma_{\xi_S}^2 \sim \text{Inverse Gamma}\ (a_{\xi_S}, b_{\xi_S}),$$

$$\sigma_{\xi_I}^2 \sim \text{Inverse Gamma}\ (a_{\xi_I}, b_{\xi_I}).$$

Notice that the uniform distributions on β and ϕ could easily be replaced by the very flexible Generalized Beta distributions on their supports.

In practice, the hyperparameter β_{max} could be data-driven using (31):

$$\beta_{max} \equiv \min_t \left(\frac{\lambda_N - Z_R(t)}{Z_S(t)Z_I(t)}\right), \tag{34}$$

where

$$Z_R(t) = \max\{1, \lambda_N - Z_S(t) - Z_I(t)\}. \qquad (35)$$

The hyperparameter μ_{θ_γ} could be data-driven through the inverse relationship between recovery rate and the mean duration of the infectious period for individuals (e.g., Lloyd, 2001), which is about 3 days for common influenzas (Centers for Disease Control;
http://www.cdc.h1n1flu/recommendations.htm). In our case, we use $\mu_{\theta_\gamma} = \log(0.33/(1 - 0.33)) = -0.708$, with $\sigma^2_{\theta_\gamma} = 0.01$. The other hyperparameters, namely, a_{ξ_S}, b_{ξ_S}, a_{ξ_I}, and b_{ξ_I}, need to be specified; for example, the choice, $a_{\xi_S} = a_{\xi_I} = 0.25$ and $b_{\xi_S} = b_{\xi_I} = 0.4$, results in a fairly vague prior for the variance components.

3 W-Scale Approximations for the HSIRS Model

Here, we derive a calibrated approximation to the nonlinear **W**-scale process in the HSIRS model. From Appendix A.2, for $t = 1, 2, ...$, equation (26) in the HSIRS model can be approximated by

$$\mathbf{W}(t + 1) = \boldsymbol{\mu}^{LW}(t) + \boldsymbol{\zeta}(t + 1), \qquad (36)$$

where recall that $\mathbf{W}(t) \equiv (W_S(t), W_I(t))'$ is the true log-odds-ratio vector.

The vector $\boldsymbol{\zeta}(t) \equiv (\zeta_S(t), \zeta_I(t))'$ in (36) is the small-scale-variation vector that captures the uncertainties in the epidemic process as well as the higher-order terms in the Taylor-series expansions. For $t = 1, 2, ...$, we assume a multivariate Normal distribution,

$$\boldsymbol{\zeta}(t) \sim \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Sigma_\zeta}(t)), \qquad (37)$$

where $\boldsymbol{\Sigma_\zeta}(t) \equiv \mathrm{diag}\left(\sigma^2_{\zeta_S}(t), \sigma^2_{\zeta_I}(t)\right)$ is the covariance matrix of $\boldsymbol{\zeta}(t)$, and $\sigma^2_{\zeta_S}(t)$ and $\sigma^2_{\zeta_I}(t)$ are nonnegative variance-component parameters. Typically, the components of $\boldsymbol{\Sigma_\zeta}(t)$ are larger than the respective components of $\boldsymbol{\Sigma_\xi}(t)$, because $\{\boldsymbol{\zeta}(t)\}$ also captures the higher-order terms left after matching the linear approximation.

In (36), the vector $\boldsymbol{\mu}^{LW}(t) \equiv (\mu_S^{LW}(t), \mu_I^{LW}(t))'$ is a linear dynamical process derived through Taylor-series expansions that approximates the nonlinear stochastic process $\boldsymbol{\mu}^W(t)$ defined in (27)-(28). From Appendix A.2, for $t = 1, 2, ...$,

$$\mu_S^{LW}(t) = J_0(t) + J_1(t)W_S(t) + J_2(t)W_I(t), \qquad (38)$$

$$\mu_I^{LW}(t) = J_3(t) + J_4(t)W_S(t) + J_5(t)W_I(t), \qquad (39)$$

where

$$J_0(t) \equiv \log\hat{A}_1(t) + \frac{1}{\hat{A}_1(t)} + \phi \left[\frac{\exp(\hat{A}_5(t))(1 - \hat{A}_5(t))}{\hat{A}_1(t)} + \frac{1}{\hat{A}_2(t)} \right] \Delta$$

$$- \frac{\beta\lambda_N}{\hat{A}_1(t)} \left(1 - \frac{1}{1 - \hat{A}_7(t)} + \frac{B_0(t) + \hat{A}_7(t)}{(1 - \hat{A}_7(t))^2} \right) \Delta$$

$$- \log\hat{A}_2(t) - \frac{1}{\hat{A}_2(t)} - \frac{\gamma}{\hat{A}_2(t)} e^{\hat{A}_6(t)}(1 - \hat{A}_6(t))\Delta, \tag{40}$$

$$J_1(t) \equiv 1 - \frac{\phi\exp\left(\hat{A}_5(t)\right)\Delta}{\hat{A}_1(t)} - \frac{\beta\lambda_N B_1(t)\Delta}{\hat{A}_1(t)(1 - \hat{A}_7(t))^2}, \tag{41}$$

$$J_2(t) \equiv -\frac{\beta\lambda_N B_2(t)\Delta}{\hat{A}_1(t)(1 - \hat{A}_7(t))^2} - \frac{\gamma e^{\hat{A}_6(t)}\Delta}{\hat{A}_2(t)}, \tag{42}$$

$$J_3(t) \equiv \log\hat{A}_3(t) + \frac{1 - \gamma\Delta}{\hat{A}_3(t)} + \frac{\beta\lambda_N\Delta}{\hat{A}_3(t)(1 - \hat{A}_{10}(t))} - \frac{\beta\lambda_N(B_3(t) + \hat{A}_{10}(t))\Delta}{\hat{A}_3(t)(1 - \hat{A}_{10}(t))^2}$$

$$- \log\hat{A}_2(t) - \frac{1}{\hat{A}_2(t)} - \frac{\gamma e^{\hat{A}_6(t)}(1 - \hat{A}_6(t))\Delta}{\hat{A}_2(t)} + \frac{\phi\Delta}{\hat{A}_2(t)}, \tag{43}$$

$$J_4(t) \equiv \frac{-\beta\lambda_N B_4(t)\Delta}{\hat{A}_3(t)(1 - \hat{A}_{10}(t))^2}, \tag{44}$$

$$J_5(t) \equiv 1 - \frac{\beta\lambda_N B_5(t)\Delta}{\hat{A}_3(t)(1 - \hat{A}_{10}(t))^2} - \frac{\gamma e^{\hat{A}_6(t)}\Delta}{\hat{A}_2(t)}. \tag{45}$$

Also,

$$B_0(t) \equiv e^{\hat{A}_4(t)}(1 - \hat{A}_4(t))B^*(t), \tag{46}$$

$$B_1(t) \equiv \frac{e^{\hat{A}_5(t) + \hat{A}_4(t)}(1 - \hat{A}_4(t))}{(1 - \hat{A}_9(t))^2} - e^{\hat{A}_4(t)}B^*(t), \tag{47}$$

$$B_2(t) \equiv e^{\hat{A}_4(t)}B^*(t), \tag{48}$$

$$B_3(t) \equiv e^{\hat{A}_4(t)}(1 - \hat{A}_4(t)) + e^{\hat{A}_5(t)}(1 - \hat{A}_5(t)), \tag{49}$$

$$B_4(t) \equiv -e^{\hat{A}_4(t)} - e^{\hat{A}_5(t)}, \tag{50}$$

$$B_5(t) \equiv e^{\hat{A}_4(t)}, \tag{51}$$

where

$$B^*(t) = \frac{1}{1 - \hat{A}_9(t)} + \frac{e^{\hat{A}_5(t)}(\hat{A}_5(t) - 1) - \hat{A}_9(t)}{(1 - \hat{A}_9(t))^2}, \tag{52}$$

and $\{\hat{A}_j(t) : j = 1, ..., 10\}$ are discussed below.

Notice that by combining equations (36)-(52), we are able to write $\mathbf{W}(t+1)$ given by (36) in a multivariate autoregressive form:

$$\mathbf{W}(t+1) = \mathbf{C}(t) + \mathbf{H}(t)\mathbf{W}(t) + \boldsymbol{\zeta}(t+1), \qquad (53)$$

where $\mathbf{C}(t) \equiv (J_0(t), J_3(t))'$ and the 2×2 matrix $\mathbf{H}(t)$ is

$$\mathbf{H}(t) \equiv \begin{pmatrix} J_1(t) & J_2(t) \\ J_4(t) & J_5(t) \end{pmatrix}.$$

The general idea behind (38)-(39) is to use $\hat{A}_i(t)$, $i = 1, ..., 10$, as an initialization of the Taylor-series expansion of the nonlinear process $\boldsymbol{\mu}^W(t)$ in (26). Formulas for $\hat{A}_i(t)$ and the quantity $A_i(t)$ that it approximates are given in Table 1. In practice, we use empirical values obtained from data $\{Z_S(t)\}$ and $\{Z_I(t)\}$ to obtain $\hat{A}_i(t)$ close to $A_i(t)$, where from Table 1 $\{A_i(t) : i = 1, ..., 10\}$ are nonlinear components in the HSIRS model. From the transformation in (24) and (25), we can obtain

$$\lambda_N P_S(t) = \frac{\lambda_N \exp(W_S(t))}{1 + \exp(W_S(t)) + \exp(W_I(t))},$$

$$\lambda_N P_I(t) = \frac{\lambda_N \exp(W_I(t))}{1 + \exp(W_S(t)) + \exp(W_I(t))},$$

$$\lambda_N P_R(t) = \frac{\lambda_N}{1 + \exp(W_S(t)) + \exp(W_I(t))}.$$

Now $\lambda_N P_S(t)$, $\lambda_N P_I(t)$, and $\lambda_N P_R(t)$ are the means of $\{Z_S(t)\}$, $\{Z_I(t)\}$, and $\{Z_R(t)\}$, and hence the values in the column "Initializations ($\hat{A}(t)$)" in Table 1 are reasonable choices for $\{\hat{A}_i(t)\}$. If data are not available at some time point, we obtain $\{\hat{A}_i(t)\}$ in Table 1 by replacing the observed data $\{Z_S(t)\}$ and $\{Z_I(t)\}$ with the predicted counts provided by a simple model, such as the CSIRS model. The goal is to obtain $\{\hat{A}_i(t)\}$ as close as possible to $\{A_i(t)\}$.

In Table 1, β_0, γ_0, and ϕ_0 are initial values of β, γ, and ϕ, respectively, which are used to enhance the Taylor-series expansions. We use values obtained from the CSIRS model (e.g., Anderson and May, 1991; Wearing et al., 2005; Burr and Chowell, 2006) for β_0, γ_0, and ϕ_0. Further details of implementation can be found in Section 4.3. We performed sensitivity studies and noticed that the MCMC based on this linear approximation is not sensitive to the initializations (even in forecasting), because the small-scale-variation vector in (36) can absorb the higher-order terms in the Taylor-series expansion (e.g., Cressie and Wikle, 2011, Section 7.3.3). Like the data model, the parameter model is unchanged, except that subscript ξ is replaced with subscript ζ; see Section 2.3 for details.

Finally, the approximate HSIRS model, which we call the *ASIRS model*, consists of the data model defined in (10)-(11), the linear dynamical process model for $\{\mathbf{W}(t)\}$ defined in (36), and the parameter model defined in (33) with ξ

replaced by ζ. Because of its computational efficiency, the ASIRS model is used in the Markov chain Monte Carlo algorithm that produces our posterior analysis.

4 Posterior Analysis of a Simulated Epidemic

4.1 Simulated Data

In this section, an epidemic is simulated from each of two processes meant to mimic an H1N1 outbreak in Franklin County, Ohio. The two datasets are from an HSIRS model (defined in Section 2) and a modified-CSIRS model (defined below), which we refer to as the HSIRS dataset and the modified-CSIRS dataset, respectively. Henceforth, the time step of the difference equations is $\Delta = 1$ day. Notice that for epidemics with a different developing time step, the value of Δ could be easily adjusted according to the infectious nature of the underlying virus.

4.1.1 The HSIRS Dataset

For simulating the HSIRS dataset from the HSIRS model, we select $\gamma^* = 0.33$ per day as the true value of γ, since there is an inverse relationship between γ and mean duration of the infectious period (the inverse relation between rate and average duration is due to the assumption that recovery is a process with constant intensity, that is, a constant rate per unit of time, which implies a negative exponential density); this is about 3 days for common influenza (Centers for Disease Control; http://www.cdc.h1n1flu/recommendations.htm). Furthermore, in epidemiology, the basic reproduction number, R_0, is defined as the number of secondary infections caused by a single infective, introduced into a population made up entirely of susceptible individuals, over this individual's course of the infection. Therefore, R_0 can be obtained by

$$R_0 = \frac{\beta \lambda_N}{\gamma}. \tag{54}$$

Typically, R_0 has a value between 1 and 2 for new strains of Influenza A in human communities (e.g., Anderson, 2006). Therefore, since $\lambda_N = 1,068,978$ in the 2000 Census in Franklin County, Ohio, and $\gamma^* = 0.33$, if we select $\beta^* = 5.1 * 10^{-7}$ as the true value of β, then we obtain $R_0 \approx 1.65$ to mimic a pandemic flu. Furthermore, for illustration, we simply assume $\phi^* = 0.05$, which is a typical value of loss of immunity for pandemic strains used in the long-term period (e.g., Bansal et al., 2010).

We now turn to the components of variance. Recall that the *signal-to-noise ratio* (SNR) can be defined as

$$\text{SNR} \equiv \frac{\mu}{\sigma}, \tag{55}$$

where μ is the signal mean and σ is the standard deviation of the noise. We denote $\text{SNR}_{W_S}(t)$ and $\text{SNR}_{W_I}(t)$ to be the SNR for the log odds ratios $W_S(t)$ and $W_I(t)$, respectively, at time t. Then, for $t = 1, 2, ..., T + F$, we have

$$\text{SNR}_{W_S}(t) \equiv \frac{\mu_S^W(t)}{\sigma_{\xi_S}(t)}, \tag{56}$$

$$\text{SNR}_{W_I}(t) \equiv \frac{\mu_I^W(t)}{\sigma_{\xi_I}(t)}, \tag{57}$$

where the right-hand sides of (56) and (57) are given by (27)-(29), and the time points $T + 1, ..., T + F$ represent a forecast period of F days.

As a baseline at time $t = 1$, we assume the rates of susceptible and infectious individuals, namely, $P_S^*(1)$ and $P_I^*(1)$ to be 0.94 and 0.01, respectively. Hence,

$$\mu_S^{W*}(1) = \log\left(\frac{0.94}{1 - 0.94 - 0.01}\right) = 2.934, \tag{58}$$

$$\mu_I^{W*}(1) = \log\left(\frac{0.01}{1 - 0.94 - 0.01}\right) = -1.609, \tag{59}$$

which results in

$$\boldsymbol{\mu}^{W*}(1) = \left(\mu_S^{W*}(1), \mu_I^{W*}(1)\right)' = (2.934, -1.609)',$$

the initial mean of the hidden log-odds-ratio vector, $\mathbf{W}(1)$.

For the simulated epidemic, we assume that the true values of SNR at time $t = 1$ are

$$\text{SNR}_{W_S}^*(1) = \text{SNR}_{W_I}^*(1) = 15.$$

More choices of SNR are discussed in Section 5. Then from (56)-(57), we have

$$\sigma_{\xi_S}^2(1) = (2.934/15)^2 = 0.038, \tag{60}$$

$$\sigma_{\xi_I}^2(1) = (-1.609/15)^2 = 0.012. \tag{61}$$

Recall that we assume $\sigma_{\xi_S}^2(t) = \sigma_{\xi_S}^2$ and $\sigma_{\xi_I}^2(t) = \sigma_{\xi_I}^2$, for $t = 1, 2, ...$. Therefore, from (60)-(61) we select $\sigma_{\xi_S}^{2*} = 0.038$, $\sigma_{\xi_I}^{2*} = 0.012$, as the true values of $\sigma_{\xi_S}^2$ and $\sigma_{\xi_I}^2$, respectively, and we write $\boldsymbol{\Sigma}_\xi^* \equiv \text{diag}(\sigma_{\xi_S}^{2*}, \sigma_{\xi_I}^{2*})$.

We simulate daily data for $T + F = 45$ days, where $T = 35$ and $F = 10$. Specifically, for $t = 1$, we simulate

$$\mathbf{W}(1) \sim \text{MVN}\left(\boldsymbol{\mu}^{W*}(1), \boldsymbol{\Sigma}_\xi^*\right); \tag{62}$$

then for $t = 2, ..., 45$, we simulate $\{\mathbf{W}(t)\}$ using (26) and obtain $\{\mathbf{P}(t)\}$ using transformations defined by (78)-(79) in Appendix A.1. Finally, we generate observed counts of susceptible and infectious individuals, $\{Z_S(t)\}$ and $\{Z_I(t)\}$, from the Poisson distribution defined in (10)-(11), conditional on $\{\mathbf{P}(t)\}$. These counts, $\{Z_S(t)\}$ and $\{Z_I(t)\}$, represent the *HSIRS dataset*.

4.1.2 The Modified-CSIRS Dataset

Before we illustrate the procedure for generating the modified-CSIRS dataset, we first define the modified-CSIRS model. Recall that the CSIRS model is deterministic, which we now modify to incorporate uncertainty in the observations and in the parameters. Specifically, the modified-CSIRS model is a hierarchical statistical model that consists of a data model defined by (10)-(11), a *deterministic* process model defined by the CSIRS model in (21)-(23), and a parameter model. That is, for $t = 1, 2, ...$, in units of Δ days, the *modified-CSIRS model* can be written as:

Data model:
$$Z_S(t)|\lambda_S(t) \sim \text{ind. Poisson}(\lambda_N P_S(t)),$$

$$Z_I(t)|\lambda_I(t) \sim \text{ind. Poisson}(\lambda_N P_I(t)).$$

Process model:
$$P_S(t+1) = P_S(t) - \beta\lambda_N P_S(t)P_I(t)\Delta + \phi P_R(t)\Delta,$$

$$P_I(t+1) = P_I(t) + \beta\lambda_N P_S(t)P_I(t)\Delta - \gamma P_I(t)\Delta,$$

$$P_R(t+1) = P_R(t) + \gamma P_I(t)\Delta - \phi P_R(t)\Delta.$$

Parameter model:
$$[\beta, \gamma, \phi] = [\beta][\gamma][\phi].$$

In this hierarchical statistical model, recall that $Z_S(t)$ and $Z_I(t)$ are the observed susceptible and infectious counts, respectively; $P_S(t)$, $P_I(t)$, and $P_R(t)$ are the hidden true proportions; β, γ, and ϕ are transmission rate, recovery rate, and loss-of-immunity rate, respectively, per day; and Δ is the time step. As specified at the beginning of this section, $\Delta = 1$. Notice that in the modified-CSIRS model, the process model is deterministic and does not capture any uncertainty in the hidden epidemic process.

For simulating the modified-CSIRS dataset from the modified-CSIRS model, we likewise simulate daily data for $T + F = 45$ days, where $T = 35$ and $F = 10$. For the unknown parameters, β, γ, and ϕ, we select the same values in Section 4.1.1, and we select the same starting proportions, $P_S^*(1) = 0.94$ and $P_I^*(1) = 0.01$. Then for $t = 2, ..., 45$, we simulate $\{\mathbf{P}(t)\}$ from the deterministic process defined in (21)-(23); and finally we generate $\{Z_S(t)\}$ and $\{Z_I(t)\}$ using the Poisson distribution defined in (10)-(11), conditional on $\{\mathbf{P}(t)\}$. These counts, $\{Z_S(t)\}$ and $\{Z_I(t)\}$, represent the *modified-CSIRS dataset*.

Notice that the HSIRS dataset mimics noisy observations from an epidemic with uncertainties in the underlying epidemic process, whereas the modified-CSIRS dataset mimics noisy observations from a deterministic epidemic process. Although for each dataset, we simulated data for $T + F = 45$ days, when we fit models, we assume that data are only available on the first $T = 35$ days and are missing in the last $F = 10$ days. Thus, we can assess the performance

of these models on forecasting, by comparing their forecasts of $\lambda_S(t)$ and $\lambda_I(t)$ to the true values known from simulation.

Figures 1(a)-1(d) show the daily observed susceptible counts $\{Z_S(t)\}$ and infectious counts $\{Z_I(t)\}$, for each of the two datasets, as a function of time (for all 45 days). We use the vertical line on each plot to emphasize that we assume data are only available for the first $T = 35$ days. Comparing these plots, we can clearly see that the epidemic patterns in the modified-CSIRS dataset (Figure 1(c)-1(d)) are much smoother than those in HSIRS dataset. In general, the two datasets suggest similar epidemic patterns in the 45 days. That is, both of them indicate clearly an epidemic between day 15 to day 40, and the infectious population reaches its peak in the period between day 20 and day 30.

4.2 Fitting the HSIRS Model

As mentioned in Section 3, we derive a well calibrated Gaussian linear process (the ASIRS model) to approximate the nonlinear $\mathbf{W}$-scale process in the HSIRS model, which improves computational efficiency in posterior analysis and forecasting. Recall that within the 45-day study period, data are only available on the first $T = 35$ days and are missing for the last $F = 10$ days.

Based on the ASIRS model specified in Section 3, the joint posterior distribution of all "unknowns" is proportional to a product of the data model, the process model, and the parameter model. Notice that $\{W_S(t)\}$ and $\{W_I(t)\}$ are transformations of $\{P_S(t)\}$ and $\{P_I(t)\}$, and hence the conditioning in the data model can be equivalently written in terms of $\mathbf{W}(t)$. Therefore, combining equations (78)-(79) in Appendix A.1, the data model given by (10) and (11) can be written as follows: For $t = 1, ..., T$,

$$Z_S(t)|\mathbf{W}(t) \sim \text{ ind. Poisson}\left(\frac{\lambda_N\exp(W_S(t))}{1 + \exp(W_I(t)) + \exp(W_S(t))}\right),$$

$$Z_I(t)|\mathbf{W}(t) \sim \text{ ind. Poisson}\left(\frac{\lambda_N\exp(W_I(t))}{1 + \exp(W_I(t)) + \exp(W_S(t))}\right).$$

Therefore, the joint posterior distribution of all unknowns is

$$[\beta, \theta_\gamma, \phi, \sigma^2_{\zeta_S}, \sigma^2_{\zeta_I}, \{W_S(t)\}, \{W_I(t)\}|Z_S(1), ..., Z_S(T), Z_I(1), ..., Z_I(T)]$$

$$\propto \prod_{t=1}^{T}[Z_S(t)|W_S(t), W_I(t)] \cdot \prod_{t=1}^{T}[Z_I(t)|W_S(t), W_I(t)] \cdot [W_S(1), W_I(1)|\sigma^2_{\zeta_S}, \sigma^2_{\zeta_I}]$$

$$\cdot \left(\prod_{t=2}^{T+F}[W_S(t), W_I(t)|\beta, \theta_\gamma, \phi, W_S(t-1), W_I(t-1), \sigma^2_{\zeta_S}, \sigma^2_{\zeta_I}]\right) \cdot [\beta][\theta_\gamma][\phi][\sigma^2_{\zeta_S}][\sigma^2_{\zeta_I}],$$

$$(63)$$

where at $t = 1$, $\mathbf{W}(1) \equiv (W_S(1), W_I(1))'$ has distribution,

$$\mathbf{W}(1)|\sigma^2_{\zeta_S}, \sigma^2_{\zeta_I} \sim \text{MVN}\left(\boldsymbol{\mu}^W(1), \boldsymbol{\Sigma}_\zeta\right),$$

and recall that $\boldsymbol{\Sigma}_\zeta = \text{diag}(\sigma^2_{\zeta_S}, \sigma^2_{\zeta_I})$. There is strong prior information on what happens at $t = 1$, which allows the hyperparameter $\boldsymbol{\mu}^W(1)$ to be specified. For example, in Section 4.1, we specify it as,

$$\boldsymbol{\mu}^W(1) = \boldsymbol{\mu}^{W*}(1) \equiv (2.934, -1.609)'.$$

(Recall from Section 4.1 that $\boldsymbol{\mu}^{W*}(1)$ is the initial mean of the log-odds-ratio vector used for simulating the data.) The parameter model defined in Section 2.3 consists of independent prior distributions on each parameter. Regarding the parameter-model specification for $\boldsymbol{\Sigma}_\zeta$, we specify fairly vague priors for $\sigma^2_{\zeta_S}$ and $\sigma^2_{\zeta_I}$ by choosing independent Inverse Gamma distributions with hyperparameters, $a_{\zeta_S} = a_{\zeta_I} = 0.25$, $b_{\zeta_S} = b_{\zeta_I} = 0.4$. Recall from Section 2.3 that we also specify $\beta \sim \text{Uniform}(0, \beta_{max})$, where β_{max} is given by (34); $\theta_\gamma \sim \text{Normal}(0.33, 0.01)$; $\phi \sim \text{Uniform}[0, 1]$.

Even based on the ASIRS model, the posterior distribution is not available analytically, owing to a normalizing constant that cannot be obtained in closed form. However, we can sample from the posterior distribution using a Markov chain Monte Carlo (MCMC) algorithm with a Gibbs sampler that incorporates Metropolis-Hastings steps where necessary (e.g., Waller et al., 1997). These are based on full conditional distributions that are given in Appendix A.3.

The MCMC algorithm draws samples, cyclically from each full conditional distribution, conditioning on the most recent samples drawn from the other full conditionals. This iterative procedure defines a Markov chain whose stationary distribution is the joint distribution of all the unknowns given the data (i.e., the posterior distribution). Hence, after a "burn-in" number of iterations, we obtain samples from the posterior distribution. Notice that in our case, except for $[\sigma^2_{\zeta_S}|\text{rest}]$ and $[\sigma^2_{\zeta_I}|\text{rest}]$, which follow the Inverse Gamma distributions, and $[W_S(t), W_I(t)|\text{rest}]$ for $t > T$, which follow a multivariate Normal distribution (in Appendix A.3, we define "rest" to represent all other unknowns as well as the data $\{Z_S(1), ..., Z_S(T)\}$ and $\{Z_I(1), ..., Z_I(T)\}$), the full conditional distributions of all the other unknowns cannot be simulated directly, and so Metropolis-Hastings updates are applied (e.g., Robert and Casella, 2004).

4.3 Posterior Analysis and Forecasting

4.3.1 Posterior Analysis Based on the HSIRS Model

For the purpose of comparison, we fit the approximate HSIRS (i.e., ASIRS) model to each of the two datasets, and we refer to them as:

− *case AH*: fit the ASIRS model to the HSIRS dataset

− *case AM*: fit the ASIRS model to the modified-CSIRS dataset

For each of these two cases, we ran an MCMC chain of 30,000 iterations. After a burn-in of 3,000 iterations, we obtained a total of 27,000 samples from the posterior distribution for each case listed above. Notice that this is *not* a particle-filtering approach in which new data are used to update current and past posteriors, so avoiding the need to re-run the MCMC.

Figure 2 shows the posterior behavior of β, γ, $\sigma^2_{\zeta_S}$, $\sigma^2_{\zeta_I}$, and ϕ, for the case AH. Similar figures are obtained for the case AM (not shown). The posterior median, mean, variance, and 95% Bayes credible interval (95%CI) for parameters in each of the two cases are shown in Table 2-3. For both cases, we can see that, except for γ (recall that θ_γ in the tables is the logit transformation of γ), the posteriors for the parameters are much tighter than the priors. See also Figure 3, which shows the prior and posterior distributions of β (chosen as a representative parameter) for both cases. Hence, there is substantial learning about these parameters, but there is almost no learning for γ because it already has a tight prior.

Because the data are simulated, we have the opportunity to compare the posterior samples with the true values. As mentioned in Section 4.1, for both datasets, we use $\beta^* = 5.1 * 10^{-7}$, $\gamma^* = 0.33$, and $\phi^* = 0.05$, as the true values of the transmission rate, the recovery rate, and the loss-of-immunity rate, respectively (in units of per day). Tables 2-3 show that ASIRS, the (approximate) hierarchical model, performs well to recover these parameters, and the posterior 95%CIs of β and ϕ are much tighter than those of their priors. (The parameter for γ already has a tight prior, and its posterior hardly changes.) Figures 2(a), 2(b), and 2(e) show the true values of the three parameters as red lines for the case AH. We find the agreements are excellent, especially for β and γ. The same conclusions hold for the case AM (not shown).

Recall that the HSIRS dataset has uncertainties associated with the hidden epidemic process, and we selected $\sigma^{2*}_{\xi_S} = 0.038$, $\sigma^{2*}_{\xi_I} = 0.012$ as the true values of the small-scale variance components. Tables 2 and 3 indicate that when fitting the ASIRS model to the HSIRS dataset (case AH), the posterior samples of the small-scale variances in the ASIRS model (i.e., $\sigma^2_{\zeta_S}$ and $\sigma^2_{\zeta_I}$) tend to be quite a bit larger than these values. These results are expected since, as explained in Section 3, the small-scale variation terms in the ASIRS model have the flexibility to capture higher-order terms not included in the linear approximations.

To assess the general performance of the ASIRS model to approximate the HSIRS model, we use a discrepancy measure, as described in Cressie and Wikle (2011), Section 2.2.2, due to Gelman et al. (1996). Based on a given discrepancy measure, we obtain a *posterior predictive p-value* and perform a diagnostic procedure to determine whether the model fits the data.

Specifically, for the kth sample from the MCMC, $k = 1, ..., m = 27,000$, we define a discrepancy measure, $\psi\left(\mathbf{Z}; \boldsymbol{\lambda}^{(k)}\right)$, as follows:

$$\psi\left(\mathbf{Z}; \boldsymbol{\lambda}^{(k)}\right) \equiv \sum_{t=1}^{T} \left[\mathbf{Z}(t) - \mathrm{E}\left(\mathbf{Z}(t)|\boldsymbol{\lambda}^{(k)}(t)\right)\right]' \left(\boldsymbol{\Sigma}_Z^{(k)}(t)\right)^{-1} \left[\mathbf{Z}(t) - \mathrm{E}\left(\mathbf{Z}(t)|\boldsymbol{\lambda}^{(k)}(t)\right)\right],$$

(64)

where $\boldsymbol{\lambda}^{(k)}(t) \equiv \left(\lambda_S^{(k)}(t), \lambda_I^{(k)}(t)\right)'$ denotes the vector of true counts at time t for the kth MCMC sample, and

$$\lambda_S^{(k)}(t) \equiv \lambda_N P_S^{(k)}(t),$$ (65)

$$\lambda_I^{(k)}(t) \equiv \lambda_N P_I^{(k)}(t).$$ (66)

Recall that we assume data are only available on the first $T = 35$ days for each of the two datasets. Based on the Poisson data model defined in (10)-(11) and on the definitions of $\boldsymbol{\lambda}^{(k)}(t)$ in (65) and (66),

$$\left(\boldsymbol{\Sigma}_Z^{(k)}(t)\right) \equiv \mathrm{diag}\left(\lambda_S^{(k)}(t), \lambda_I^{(k)}(t)\right),$$

represents the data model's covariance matrix, and hence (64) take the form of a Wald statistic. Furthermore, for $t = 1, ..., T$, if we use $\mathbf{Z}_{rep}(t)$ to denote an independent replicate of the data, then the posterior-predictive distribution of $\mathbf{Z}_{rep}(t)$ can be defined as (Gelman et al., 1996),

$$[\mathbf{Z}_{rep}(t)|\mathbf{Z}(t)] = \int \int [\mathbf{Z}_{rep}(t)|\{\mathbf{W}(t)\}, \boldsymbol{\Theta}] \, [\{\mathbf{W}(t)\}, \boldsymbol{\Theta}|\mathbf{Z}(t)] \, d\{\mathbf{W}(t)\} d\boldsymbol{\Theta}(t),$$

(67)

where

$$\boldsymbol{\Theta} \equiv \{\beta, \gamma, \phi, \sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2\}.$$

Thus, for the kth sample, $\mathbf{Z}_{rep}^{(k)}(t)$ is drawn from the posterior distribution $[\mathbf{Z}_{rep}(t)|\mathbf{Z}(t)]$, $t = 1, ..., T$, and we obtain the discrepancy measure $\psi\left(\mathbf{Z}_{rep}^{(k)}; \boldsymbol{\lambda}^{(k)}\right)$ by replacing $\mathbf{Z}$ in (64) with $\mathbf{Z}_{rep}^{(k)}(t)$ defined in (67). Therefore, the replicates,

$$\{\mathbf{Z}_{rep}^{(1)}(t), \mathbf{Z}_{rep}^{(2)}(t), ..., \mathbf{Z}_{rep}^{(m)}\},$$

should "look like" the data $\mathbf{Z}(t)$ if the model is appropriate. Based on this idea, we can apply posterior-predictive diagnostics in the case AH and the case AM. Since the HSIRS dataset is simulated from the HSIRS model, we can also assess the performances of the ASIRS model to approximate the HSIRS model.

The posterior predictive p-value can be obtained as below (Gelman et al., 1996):

$$\Pr\left(\psi(\mathbf{Z}_{rep}; \boldsymbol{\lambda}) \geq \psi(\mathbf{Z}; \boldsymbol{\lambda})\right) = \frac{1}{m} \sum_{k=1}^{m} \mathrm{I}\left[\psi\left(\mathbf{Z}_{rep}^{(k)}; \boldsymbol{\lambda}^{(k)}\right) \geq \psi\left(\mathbf{Z}; \boldsymbol{\lambda}^{(k)}\right)\right], \quad (68)$$

where I($\cdot$) denotes an indicator function.

Based on (68), we obtain the posterior predictive p-values for the case AH, to be 0.458; hence, there is no striking evidence for a lack of model fit when fitting the ASIRS model to the HSIRS dataset. This conclusion is further supported by Figure 4, which plots $\psi\left(\mathbf{Z}_{rep}^{(k)}; \boldsymbol{\lambda}^{(k)}\right)$ against $\psi\left(\mathbf{Z}; \boldsymbol{\lambda}^{(k)}\right)$ for the case AH; there is no striking departure from the 45-degree line. Indeed, the same is true for the case AM, and we conclude that there is no evidence for lack of fit for the ASIRS model when fitted to either dataset. That is, the ASIRS model appears to have considerable flexibility.

4.3.2 Comparisons of the fitted ASIRS model and the fitted CSIRS model

We now fit the (discrete-time) CSIRS model to each of the two simulated datasets described in Section 4.1. In practice, estimates of β, γ, and ϕ, denoted as $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\phi}$, are needed in order to solve for $\{S(t)\}$ and $\{I(t)\}$ in equations (6)-(8). Here, we minimize the sum of squares between the estimated and observed infectious counts over time (e.g., Anderson and May, 1991; Wearing et al., 2005; Burr and Chowell, 2006). That is,

$$(\hat{\beta}, \hat{\gamma}, \hat{\phi}) = \arg \min_{(\beta, \gamma, \phi)} \left[\sum_{t=1}^{T} \left(\hat{I}(t; \beta, \gamma, \phi) - Z_I(t) \right)^2 \right], \qquad (69)$$

where $\hat{I}(t; \beta, \gamma, \phi)$ is the deterministic estimate of infectious counts at time t, obtained after substituting β, γ, and ϕ into equations (6)-(8). Recall that here $T = 35$, and also notice that these least squares estimates, $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\phi}$, are the values selected in Section 3 for the initial MCMC values (denoted β_0, γ_0, and ϕ_0, there.) Finally, the CSIRS forecasts are given by (6)-(8) for $t = T + 1, ..., T+F$; using obvious notation, they are $\{\hat{S}(t; \hat{\beta}, \hat{\gamma}, \hat{\phi}) : t = T+1, ..., T+F\}$, $\{\hat{I}(t; \hat{\beta}, \hat{\gamma}, \hat{\phi}) : t = T + 1, ..., T + F\}$, and $\{\hat{R}(t; \hat{\beta}, \hat{\gamma}, \hat{\phi}) : t = T + 1, ..., T + F\}$, where $T = 35$ and $F = 10$. For $t = 1, ..., T$, the observed counts are considered to be the CSIRS-predicted counts.

Now, for the HSIRS model (approximated by ASIRS), the appropriate predictions are obtained from posterior inference on $\lambda_S(t)$, and $\lambda_I(t)$, for the entire study period $t = 1, ..., T, T + 1, ..., T + F$, but based only on data from $t = 1, ..., T$. The Bayesian predictions are actually given for $P_S(t)$ and $P_I(t)$; then to obtain ASIRS-predicted counts, equations (13)-(15) along with mass balance given by (17) are used.

In the infectious-disease setting, the number of infectious individuals is one of the most important quantities of interest. The estimated trajectories given in Figure 5 show infectious counts for the entire 45-day study period, obtained by fitting the ASIRS model and the CSIRS model to each of the two datasets. These are compared to the true (hidden) values of infectious counts. Consider the modified-CSIRS dataset: upon inspection of Figure 5(b), we see that both

models can capture the overall epidemic pattern very well, not only during the first 35 days when data are available, but also in the last 10 days when there are no data. Now consider the HSIRS dataset: when the underlying epidemic process has stochastic components, the disadvantage of the CSIRS model becomes apparent; upon inspection of Figure 5(a), estimates from fitting the CSIRS model are oversmoothed, even on the days when data are available. In contrast, the agreement between the true value and the posterior median of $\{\lambda_I(t)\}$, obtained from fitting the ASIRS model, is excellent when data are available. On the last 10 days when there are no data, the hierarchical model is able to predict the general downward trend at the end of the epidemic process; however, the CSIRS model is unable to deal with the uncertainties, mistakenly forecasting that the epidemic is maintained in the last 10 days.

Another disadvantage of the CSIRS model is that it is unable to provide any uncertainty measures to accompany its deterministic-modeling strategy. In contrast, when fitting the hierarchical model, we can obtain uncertainty measures for any quantity of interest, based on its posterior distribution. For example, Figure 6 shows 0.025, 0.25, 0.5, 0.75, 0.975 quantiles of posterior distributions of the hidden infectious counts, $\lambda_I(t)$, during the forecasting period $\{T + 1, ..., T + F\}$. We can see that in all cases, the posterior 50%CI of $\lambda_I(t)$ obtained from fitting the ASIRS model, cover the true values at all times.

Now consider the susceptible counts. Figure 7 and Figure 8 clearly indicate that the ASIRS model performs better than the CSIRS model, when fitting to the HSIRS dataset.

The analyses given in this section illustrate the advantages and the practicalities associated with a hierarchical statistical approach to inferring the dynamical evolution of an infectious disease. In the next section, we confirm those inferential advantages through a simulation experiment.

5 A Simulation Experiment

In this section, a simulation experiment is presented to compare the performances of the HSIRS model (actually, the approximative ASIRS model) and the CSIRS model using design-based criteria under various factor combinations (i.e., a factorial experimental design), assuming that data are from an HSIRS model (defined in Section 2). That is, the processes $\{\mathbf{W}(t)\}$ and $\{\mathbf{P}(t)\}$ are simulated according to (26) and (78)-(79), from which the data processes $\{Z_S(t)\}$ and $\{Z_I(t)\}$ are simulated from the Poisson distributions defined in (10)-(11), conditional on $\{\mathbf{P}(t)\}$. As in Section 4.1.1, we set the true values of the parameters in (26) as ($\gamma^* = 0.33$, $\phi^* = 0.05$), and we simulate the baseline $\mathbf{W}(1)$ from (62) with $\boldsymbol{\mu}^{W*}(1) = (2.934, -1.609)$. The baseline small-scale variation $\boldsymbol{\Sigma}_\xi^*$ in (62), the values of T and F, and the parameter β in (26) are chosen in ways that relate to factors in the experiment and will be described

in Section 5.1. For each of the factor combinations, we simulate $L = 100$ replications.

5.1 Factors of the Simulation Experiment

Four factors are considered in this simulation experiment: fitted model (**FM**) is a factor that compares the HSIRS model and the CSIRS model; in the terminology of experimental design, this is considered the "treatment." Three other factors related to the data-generating schemes are also included; these are data information (**DI**), small-scale variation (**SV**), and transmission rate (**TR**). The details of these four factors are now presented.

Fitted Model (FM)
The factor **FM** has two levels, where **FM**=0 represents the CSIRS model and **FM**=1 represents the HSIRS model. The HSIRS model is specified by (10), (11), and (26) in Section 2, and by (78), (79), and (80) in Appendix A.1. The CSIRS model is specified by (6)-(8) in Section 1.

Data Information (DI)

Two different levels of **DI** are considered. One is where there are data for 15 days (**DI**=0), and the other for 35 days (**DI**=1). In each situation, we perform forecasting for up to 10 days beyond the last day where there were data. The choice of 15 days (**DI**=0) was based on CDC's response to the 2009 H1N1 pandemic. According to "The 2009 H1N1 Pandemic: Summary Highlights, April 2009-April 2010" (CDC: http://www.cdc.gov/h1n1flu/cdcresponse.html), an H1N1 infection was first detected in the US in a 10-year-old patient in California on April 15, 2009; 15 days later, that is, on April 29, 2009, WHO raised the influenza pandemic alert from phase 4 to phase 5, signaling that a pandemic was imminent. We selected 35 days (**DI**=1) because, in retrospect, the 2009 H1N1 pandemic took 35 days to develop from the time the first case occurred in Mexico (Figure 9, from "Outbreak of Swine-Origin Influenza A (H1N1) Virus Infection—Mexico, March–April 2009";
CDC: http://www.cdc.gov/mmwR/preview/mmwrhtml/mm58d0430a2.html).

Small-scale Variation (SV)

Two different levels of **SV** are considered. Recall the definitions of signal-to-noise ratios, $\mathrm{SNR}_{W_S}(t)$ and $\mathrm{SNR}_{W_I}(t)$, in equations (56) and (57) in Section 4.1.1. These are put equal to each other, and the two levels are SNR=15 (**SV**=0) and SNR=5 (**SV**=1). Consider the initial time $t = 1$. Then for **SV**=0,

$$\mathrm{SNR}_{W_S}(1) = \mathrm{SNR}_{W_I}(1) = 15;$$

hence (60)-(61) gives $\sigma^2_{\xi_S}(1) = 0.038$, $\sigma^2_{\xi_I}(1) = 0.012$. For **SV** $= 1$,

$$\mathrm{SNR}_{W_S}(1) = \mathrm{SNR}_{W_I}(1) = 5;$$

then from (56)-(57), we have

$$\sigma^2_{\xi_S}(1) = (2.934/5)^2 = 0.344,$$

$$\sigma^2_{\xi_I}(1) = (-1.609/5)^2 = 0.104.$$

To control this factor in the simulation, we assume henceforth that $\sigma^2_{\xi_S}(t) \equiv \sigma^2_{\xi_S}(1)$ and $\sigma^2_{\xi_I}(t) \equiv \sigma^2_{\xi_I}(1)$, for $t = 1, 2, ...$ Therefore, **SV**=0 represents $\{\sigma^2_{\xi_S}(t) = 0.038, \sigma^2_{\xi_I}(t) = 0.012, \text{SNR}=15\}$, and **SV**=1 represents $\{\sigma^2_{\xi_S}(t) = 0.344, \sigma^2_{\xi_I}(t) = 0.104, \text{SNR}=5\}$.

Transmission Rate (TR)

Two levels of the transmission rate (**TR**) per unit time are considered: $\beta = 3.3958 \times 10^{-7}$ (**TR**=0) and $\beta = 5.8654 \times 10^{-7}$ (**TR**=1). These two levels of β result in the basic reproduction number $R_0 = 1.1$ and 1.9, respectively, given by equation (54). According to Anderson (2006), R_0 usually has a value between 1 and 2 for new strains of influenza in the human community. Therefore, by selecting these two levels of **TR**, we are aiming to mimic scenarios with small and large transmission rates, respectively, but still within the usual range.

5.2 Results of the Simulation Experiment

In order to compare the approximate HSIRS model and the CSIRS model, we define a response variable based on the *empirical mean squared prediction error* (MSPE) in each simulation run: Let T^O denote the times at which data are observed, and let T^M denote the times at which data are missing; note that in our case, T^M is a 10-day forecast period that follows T^O. Let $P_X(t, l)$ denote the l-th simulated realization of the underlying true *rate* of the population in subgroup X (see equations (13)-(15) for the definition of "true rate"); in our case, X denotes either the susceptible (S) or the infectious (I) population. Let $\hat{P}_X(t, l)$ denote a generic predictor of $P_X(t, l)$. Note that $P_X(t, l)$ contains the same information as $\lambda_X(t, l)$, which denotes the l-th realization of the simulated true counts for the *population* in subgroup X, since $\lambda_X(t, l) = \lambda_N P_X(t, l)$ and λ_N is a known constant total population. Then we define

$$MSPE_X(T^*, l) \equiv \frac{1}{|T^*|} \sum_{t \in T^*} \left(\hat{P}_X(t, l) - P_X(t, l) \right)^2 ; \ l = 1, ..., L, \qquad (70)$$

where $T^* = T^O$ or T^M, $|T^*|$ is the number of days in the time period T^*, and L is the total number of simulation runs for each of the factor combinations in this study. The bias of the predictor can be studied through,

$$BIAS_X(T^*, l) \equiv \frac{1}{|T^*|} \sum_{t \in T^*} \left(\hat{P}_X(t, l) - P_X(t, l) \right) ; \ l = 1, ..., L. \qquad (71)$$

We performed an analysis of variance (ANOVA) to investigate which factors are important and under which scenarios the approximate HSIRS model provides substantial improvement over the CSIRS model. First, we consider MSPE. Because of the skewness and mean-variance dependence in $\{MSPE_X(T^*, l)\}$, we use a fourth-root transformation to transform the response (Cressie et al., 2010). The boxplots in Figures 10 and 11 show that the distributions of the differences between the fourth-root transformations (superscripts "C" and "H" denote, respectively, analysis under the CSIRS model and under the approximate HSIRS model):

$$\left\{ MSPE_X^C(T^*, l)^{1/4} - MSPE_X^H(T^*, l)^{1/4} : l = 1, ..., L; \; X = S, I; \; T^* = T^O, T^M \right\},$$

is suitable for a classical ANOVA. We then define

$$A_X^Y(T^*) \equiv \text{ave}_l \left\{ MSPE_X(T^*, l)^{1/4} \right\}; \; X = S, I; Y = C, H; \; T^* = T^O, T^M, \tag{72}$$

where the average is taken over $L = 100$ replications for each of the factor combinations. We construct the "response" variable of the ANOVA study as $A_X^C(T^*) - A_X^H(T^*)$, for $X = S$ and I. These *paired comparisons* will tell us for which factor combinations the approximate HSIRS model shows improvement over the CSIRS model. Four ANOVAs showing up to two-way interactions are reported in Tables 4–7; they are based on the response variables, $A_S^C(T^O) - A_S^H(T^O)$, $A_S^C(T^M) - A_S^H(T^M)$, $A_I^C(T^O) - A_I^H(T^O)$, and $A_I^C(T^M) - A_I^H(T^M)$, respectively.

Table 4 gives the paired-comparison ANOVA for the infectious population, for $T^* = T^O$. It indicates that the main effect of the small-scale variation factor, **SV**, explains over 65% of the variability in the response. We also find that the difference, $A_S^C(T^O) - A_S^H(T^O)$, increases by 83% when **SV**=1, relative to its value when **SV**=0. The remaining variability is mostly explained by the main effects of **DI** and **TR**. Figure 12 supports the results shown in Table 4, where the red and blue lines show that the **SV-TR**, **SV-DE**, and **TR-DI** interactions for $A_S^C(T^O) - A_S^H(T^O)$ are not obvious. Moreover, from Figure 12, we see that for all factor combinations, the approximate HSIRS model always outperforms the CSIRS model, for $T^* = T^O$, since the responses are always greater than zero. Notice that they increase as data information becomes longer (**DI**=1), as SNR becomes smaller (**SV**=1), and as the transmission rate becomes larger (**TR**=1), which are all reasonable results that match our intuition. Clearly, as more variability is introduced into the pandemic and its observed counts, the relative inability of the CSIRS model to handle it becomes more obvious.

Table 5 gives the paired-comparison ANOVA for the infectious population, again for $T^* = T^O$. It indicates that **SV** and **TR** are the two most important factors; their main effects together explain over 89% of the variability in the response. By investigating the main effects of **SV** and **TR**, we find that

the difference, $A_I^C(T^O) - A_I^H(T^O)$, increases by 85% when **SV**=1, relative to its value at **SV**=0; and it increases by 81% when **TR**=1, relative to its value at **TR**=0. The pattern is broadly consistent with that for the susceptible population, and the responses increase when data information become longer (**DI**=1), as we expect. These results are supported further by Figure 13, where the red and blue lines show the **SV-TR**, **SV-DI**, and **TR-DI** interactions for $A_I^C(T^O) - A_I^H(T^O)$; it again indicates no evident interactions. Importantly, Figure 13 shows that for $T^* = T^O$ and all factor combinations, the approximate HSIRS model outperforms the CSIRS model.

Forecasting in the period T^M is of more interest, so we turn our attention to the susceptible populations, Table 6, and the ANOVA of $A_S^C(T^M) - A_S^H(T^M)$. It indicates that **DI** is the most dominant, which explains over 60% of the variability in the response. We also find that the difference, $A_S^C(T^M) - A_S^H(T^M)$, increases by 145% when **DI**=1, relative to its value when **DI**=0. The remaining variability is mostly explained by the main effect of **SV** and its interaction with **TR**. The green and black lines in Figure 12 show the **SV-TR**, **SV-DE**, and **TR-DI** interactions for $A_S^C(T^M) - A_S^H(T^M)$. It supports the results shown in Table 6, namely that the **SV-TR** interaction is more evident than the other two interactions. We see that when **TR**=0, the value of $A_S^C(T^M) - A_S^H(T^M)$ increases substantially when the SNR becomes smaller (**SV**=1), in contrast to very little change between the two levels of **SV** when **TR**=1. From Figure 12, the values of $A_S^C(T^M) - A_S^H(T^M)$ are always greater than zero, and they increase as SNR becomes smaller (**SV**=1) or as data information becomes longer (**DI**=1), which matches our intuition. This again indicates that for the susceptible population, the approximate HSIRS model always outperforms the CSIRS model holding all other factors the same, even though the difference is not as pronounced as during the period T^O. Since forecasting is inherently difficult, this is not surprising.

Table 7 gives the paired-comparison ANOVA for the infectious population in the period T^M: the **SV-TR** interaction has the largest impact on the response and explains about 50% of the variability. The main effects of **TR**, **SV**, and the **SV-DI** interaction explain most of the remaining variation. These results are supported further by Figure 13, where the green and black lines show the **SV-TR**, **SV-DI**, and **TR-DI** interactions for $A_I^C(T^M) - A_I^H(T^M)$. We see that the **SV-TR** interaction is similar to that for the susceptible population, but it is more evident, because when **TR**=1, the value of $A_I^C(T^M) - A_I^H(T^M)$ decreases as SNR becomes smaller (**SV**=1). A similar pattern is seen for the **DI-SV** interaction; however, the **DI-TR** interaction is not as evident as the other two interactions. Notice that at both levels of **DI**, the values of $A_I^C(T^M) - A_I^H(T^M)$ are larger when **TR** is smaller (**TR**=0). Importantly, Figure 13 shows that for $T^* = T^M$ and all factor combinations, the values of $A_I^C(T^M) - A_I^H(T^M)$ are always greater than zero; that is, the approximate HSIRS model always outperforms the CSIRS model.

To investigate further the performances of the approximate HSIRS model and the CSIRS model, we turn our attention to bias (as specified in (71)) and define the fraction of MSPE that can be explained by the squared bias as

$$\kappa_X^Y(T^*, l) \equiv \frac{BIAS_X^Y(T^*, l)^2}{MSPE_X^Y(T^*, l)}; \ X = S, I; \ T^* = T^O, T^M; \ Y = H, C; \ l = 1, ..., L.$$
(73)

Notice that $0 \le \kappa_X^Y(T^*, l) \le 1$. The boxplots in Figures 14 and 15 show the distributions of $\{\kappa_S^H(T^*, l)\}$ and $\{\kappa_I^H(T^*, l)\}$, respectively, for all factor combinations. For both the susceptible and infectious populations, Figures 14 and 15 show that the fractions of MSPE explained by the squared bias are much smaller (around 0.1 or less) in the period T^O than those in the period T^M (around 0.5), for all factor combinations. Again because forecasting is inherently difficult, this is not surprising. By comparing Figures 14 and 15 with Figures 10 and 11, we see that those factor combinations where squared bias is relatively small are the same combinations where the approximate HSIRS model yields much better MSPE than that for the CSIRS model. In particular, Figures 14 and 15 clearly show that in the period of T^O, the responses decrease (or do not increase) as data information becomes longer (**DI**=1), as SNR becomes smaller (**SV**=1), and as the transmission rate becomes larger (**TR**=1), which supports our conclusion that the approximate HSIRS model has better predictive capability even when more variability is introduced into the pandemic through its observed counts. Figures 14 and 15 indicate the strong interactions among **DI**, **SV**, and **TR** in the period of T^M, which agrees with the conclusions drawn from Figures 12 and 13.

6 Discussion and Conclusions

In this article, we develop a Bayesian hierarchical SIRS (HSIRS) model that captures the various sources of uncertainties in modeling infectious diseases such as seasonal or pandemic influenza. Important features of our HSIRS model are that it preserves mass balance on the (hidden) true counts rather than on the observed counts, and that the dynamical process is modeled on a log-odds-ratio scale. Furthermore, our approach captures the stochastic and discrete nature of the epidemic process, as well as keeping the SIRS flow that underlies the classic SIRS (CSIRS) model.

In Section 4, we simulated two datasets, an HSIRS dataset and a modified-CSIRS dataset, where we assumed that data were available on the first 35 days and missing on the following 10 days. Then we used an MCMC algorithm to fit the HSIRS model to each of the two datasets; for computational efficiency, a well calibrated linear approximation was used. We saw that the approximate HSIRS (ASIRS) model was a very good approximation, and that accounting for all known uncertainties led to a superior performance over the deterministic classic SIRS (CSIRS) model.

In Section 5, a carefully designed simulation experiment at various levels of various factors with sufficient replication is presented. It allows us to conclude that the approximate HSIRS model offers an accurate and computationally efficient approach to analyzing infectious-disease data. The comparisons given there clearly show that the HSIRS model is better, according to both the MSPE and bias criteria, than the CSIRS model, during either the observation period or the forecast period for all the conditions tested.

In this research, the HSIRS model is based on assuming a constant population for a short period of time; however, the model is not restricted to this assumption and can be extended to include population turnover that handles long-term influenza dynamics. For example, equations (21)-(23) in Section 2.2 could instead be derived from an SIRS flow with birth and death rates (see the SIR flow incorporating birth and death rates given in Anderson and May, 1991). This extension is useful because recovery from seasonal influenza may give immunity to that specific pathogen, but this immunity can be lost over the years due to the evolution of the virus (e.g., Dushoff et al., 2004). In ongoing research, we are investigating more complicated epidemic dynamics that not only incorporate birth, death, but also emigration/immigration processes for appropriate time periods. Also, spatial (e.g., Hooten and Wikle, 2010; Oleson and Wikle, 2013) and multivariate (e.g., Zhuang et al., 2013) aspects could be incorporated into these hierarchical dynamical models through vector-valued processes, although the form of such models would require careful integration of the aforementioned emigration/immigration processes.

Acknowledgments

Appendix

A.1

Here we derive the nonlinear dynamical structure of $\boldsymbol{\mu}^W(t)$ given in (27) and (28) in Section 2.2.

Assume $P_R(t) > 0$, for $t = 1, 2, \ldots$. From the difference equations (21)-(23), we can obtain

$$\frac{P_S(t+1)}{P_R(t+1)} = \frac{P_S(t) - \beta\Delta\lambda_N P_S(t)P_I(t) + \phi\Delta P_R(t)}{P_R(t) + \gamma\Delta P_I(t) - \phi\Delta P_R(t)}, \tag{74}$$

$$\frac{P_I(t+1)}{P_R(t+1)} = \frac{P_I(t) + \beta\Delta\lambda_N P_S(t)P_I(t) - \gamma\Delta P_I(t)}{P_R(t) + \gamma\Delta P_I(t) - \phi\Delta P_R(t)}. \tag{75}$$

Notice that equations (74)-(75) can be rewritten as

$$\frac{P_S(t+1)}{P_R(t+1)} = \frac{\frac{P_S(t)}{P_R(t)}\frac{1}{P_R(t)} - \beta\Delta\lambda_N\frac{P_S(t)}{P_R(t)}\frac{P_I(t)}{P_R(t)} + \frac{\phi\Delta}{P_R(t)}}{\frac{1}{P_R(t)} + \gamma\Delta\frac{P_I(t)}{P_R(t)}\frac{1}{P_R(t)} - \frac{\phi\Delta}{P_R(t)}}, \tag{76}$$

$$\frac{P_I(t+1)}{P_R(t+1)} = \frac{\beta\Delta\frac{P_S(t)}{P_R(t)}\frac{P_I(t)}{P_R(t)} + (1 - \gamma\Delta)\frac{P_I(t)}{P_R(t)}\frac{1}{P_R(t)}}{\frac{1}{P_R(t)} + \gamma\Delta\frac{P_I(t)}{P_R(t)}\frac{1}{P_R(t)} - \frac{\phi\Delta}{P_R(t)}}. \tag{77}$$

From (24)-(25),

$$P_S(t) = \frac{\exp(W_S(t))}{1 + \exp(W_S(t)) + \exp(W_I(t))}, \tag{78}$$

$$P_I(t) = \frac{\exp(W_I(t))}{1 + \exp(W_S(t)) + \exp(W_I(t))}. \tag{79}$$

Then substituting (78)-(79) into (17), we obtain

$$P_R(t) = \frac{1}{1 + \exp(W_S(t)) + \exp(W_I(t))}. \tag{80}$$

Hence,

$$\frac{P_S(t)}{P_R(t)} = \exp(W_S(t)), \tag{81}$$

$$\frac{P_I(t)}{P_R(t)} = \exp(W_I(t)) \tag{82}$$

$$\frac{1}{P_R(t)} = \frac{\exp(W_S(t)) + \exp(W_I(t)) + 1}{1}. \tag{83}$$

For $t = 1, 2, ...$, substitute (81)-(83) into equations (76)-(77) to obtain

$$
\begin{aligned}
\exp(W_S(t+1)) = \exp\left(W_S(t)\right) \\
\cdot \left[1 + \frac{\phi\Delta}{\exp\left(W_S(t)\right)} - \frac{(\beta\Delta\lambda_N)\exp\left(W_I(t)\right)}{1 + \exp\left(W_S(t)\right) + \exp\left(W_I(t)\right)}\right] \\
\cdot \frac{1}{\left[1 + \gamma\Delta\exp\left(W_I(t)\right) - \phi\Delta\right]},
\end{aligned}
\tag{84}
$$

$$
\begin{aligned}
\exp\left(W_I(t+1)\right) = \exp\left(W_I(t)\right) \\
\cdot \left[1 - \gamma\Delta + \frac{(\beta\Delta\lambda_N)\exp\left(W_S(t)\right)}{1 + \exp\left(W_S(t)\right) + \exp\left(W_I(t)\right)}\right] \\
\cdot \frac{1}{\left[1 + \gamma\Delta\exp\left(W_I(t)\right) - \phi\Delta\right]}.
\end{aligned}
\tag{85}
$$

Taking logrithms on both sides of (84)-(85), for $t = 1, 2, ...$, we obtain,

$$
\begin{aligned}
W_S(t+1) = W_S(t) \\
+ \log\left[1 + \frac{\phi\Delta}{\exp\left(W_S(t)\right)} - \frac{\beta\Delta\lambda_N\exp\left(W_I(t)\right)}{1 + \exp\left(W_S(t)\right) + \exp\left(W_I(t)\right)}\right] \\
+ \log\left[\frac{1}{1 + \gamma\Delta\exp\left(W_I(t)\right) - \phi\Delta}\right],
\end{aligned}
\tag{86}
$$

$$
\begin{aligned}
W_I(t+1) = W_I(t) \\
+ \log\left[1 - \gamma\Delta + \frac{\beta\Delta\lambda_N\exp\left(W_S(t)\right)}{1 + \exp\left(W_S(t)\right) + \exp\left(W_I(t)\right)}\right] \\
+ \log\left[\frac{1}{1 + \gamma\Delta\exp\left(W_I(t)\right) - \phi\Delta}\right].
\end{aligned}
\tag{87}
$$

Then (86)-(87) are used to define $\boldsymbol{\mu}^W(t)$ in the nonlinear autoregressive structure given by (26), which captures the uncertainties in the hidden epidemic process.

A.2

Here, we derive a well calibrated linear process $\{\boldsymbol{\mu}^{LW}(t)\}$, as defined by (38)-(39), to approximate $\{\boldsymbol{\mu}^W(t)\}$ given by (27)-(28). Specifically, for $t = 1, 2, ...$, $\{\hat{A}_i(t) : i = 1, ..., 10\}$ are initializations of the nonlinear components in equations (27)-(28), as shown in Table 1.

Consider $\mu_S^W(t)$ given by (27), and use a Taylor-series expansion up to second order. The second term on the right-hand side is:

$$\log\left(1 + \frac{\phi\Delta}{e^{W_S(t)}} - \frac{\beta\Delta\lambda_N e^{W_I(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}}\right)$$

$$=\log(A_1(t)) + \log\left[1 + \left(\frac{1 + \frac{\phi\Delta}{e^{W_S(t)}} - \frac{\beta\Delta\lambda_N e^{W_I(t)}}{\left(1 + e^{W_S(t)} + e^{W_I(t)}\right)}}{A_1(t)} - 1\right)\right]$$

$$\approx\log(\hat{A}_1(t)) + \left(\frac{1 + \frac{\phi\Delta}{e^{W_S(t)}} - \frac{\beta\Delta\lambda_N e^{W_I(t)}}{\left(1 + e^{W_S(t)} + e^{W_I(t)}\right)}}{\hat{A}_1(t)} - 1\right)$$

$$-\frac{1}{2}\left(\frac{1 + \frac{\phi\Delta}{e^{W_S(t)}} - \frac{\beta\Delta\lambda_N e^{W_I(t)}}{\left(1 + e^{W_S(t)} + e^{W_I(t)}\right)}}{\hat{A}_1(t)} - 1\right)^2. \tag{88}$$

Now, $\frac{e^{W_I(t)}}{\left(1 + e^{W_S(t)} + e^{W_I(t)}\right)}$ can be further expanded using a Taylor series to second order:

$$\frac{e^{W_I(t)}}{\left(1 + e^{W_S(t)} + e^{W_I(t)}\right)} = 1 - \frac{1}{1 - \left(-\frac{e^{W_I(t)}}{1 + e^{W_S(t)}}\right)}$$

$$\approx 1 - \left[\frac{1}{1 - \hat{A}_7(t)} + \frac{-\frac{e^{W_I(t)}}{1 + e^{W_S(t)}} - \hat{A}_7(t)}{(1 - \hat{A}_7(t))^2} + \frac{\left(-\frac{e^{W_I(t)}}{1 + e^{W_S(t)}} - \hat{A}_7(t)\right)^2}{(1 - \hat{A}_7(t))^3}\right]. \tag{89}$$

Then we expand the remaining nonlinear component in (89), $\frac{e^{W_I(t)}}{1 + e^{W_S(t)}}$, using a Taylor series to second order:

$$\frac{e^{W_I(t)}}{1 + e^{W_S(t)}} = \left[e^{(W_I(t) - W_S(t))}\right] \cdot \left[\frac{1}{1 - (-e^{-W_S(t)})}\right]$$

$$\approx \left[e^{\hat{A}_4(t)} + e^{\hat{A}_4(t)}(W_I(t) - W_S(t) - \hat{A}_4(t)) + \frac{e^{\hat{A}_4(t)}}{2}(W_I(t) - W_S(t) - \hat{A}_4(t))^2\right]$$

$$\cdot \left[\frac{1}{1 - \hat{A}_9(t)} + \frac{\left(-e^{-W_S(t)} - \hat{A}_9(t)\right)}{\left(1 - \hat{A}_9(t)\right)^2} + \frac{-e^{W_S(t)} - \hat{A}_9(t)}{\left(1 - \hat{A}_9(t)\right)^3}\right], \tag{90}$$

and

$$e^{-W_S(t)} \approx e^{\hat{A}_5(t)} + e^{\hat{A}_5(t)}\left(-W_S(t) - \hat{A}_5(t)\right) + \frac{e^{\hat{A}_5(t)}\left(-W_S(t) - \hat{A}_5(t)\right)^2}{2}. \tag{91}$$

Upon substituting (91) into (90), we obtain:

$$
\begin{aligned}
\frac{e^{W_I(t)}}{1 + e^{W_S(t)}} = & B_0(t) + B_1(t)W_S(t) + B_2(t)W_I(t) \\
& + o(W_S(t)^2) + o(W_I(t)^2) + o\left(W_S(t)W_I(t)\right),
\end{aligned}
\tag{92}
$$

where $\{B_i(t) : i = 0, 1, 2\}$ are defined in equations (46)-(48) and (52).

Hence, combining (88), (89), and (92), we can finally approximate

$$
\log\left[1 + \frac{\phi\Delta}{\exp\left(W_S(t)\right)} - \frac{\beta\Delta\lambda_N e^{W_I(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}}\right]
$$

in equation (27) with,

$$
\begin{aligned}
& \log\hat{A}_1(t) + \frac{1}{\hat{A}_1(t)} + \phi\Delta\left[\frac{e^{(\hat{A}_5(t))}(1 - \hat{A}_5(t))}{\hat{A}_1(t)}\right] \\
& - \frac{\beta\Delta\lambda_N}{\hat{A}_1(t)}\left(1 - \frac{1}{1 - \hat{A}_7(t)} + \frac{B_0(t) + \hat{A}_7(t)}{(1 - \hat{A}_7(t))^2}\right) - 1 \\
& + \left[1 - \frac{\phi\Delta e^{\hat{A}_5(t)}}{\hat{A}_1(t)} - \frac{\beta\Delta\lambda_N B_1(t)}{\hat{A}_1(t)(1 - \hat{A}_7(t))^2}\right]W_S(t) - \left[\frac{\beta\Delta\lambda_N B_2(t)}{\hat{A}_1(t)(1 - \hat{A}_7(t))^2}\right]W_I(t).
\end{aligned}
\tag{93}
$$

Also, the third term, $\log\left(1 + \Delta\gamma e^{W_I(t)} - \Delta\phi\right)$, on the right-hand side of (27), can be expanded using a Taylor series to second order:

$$
\begin{aligned}
\log\left(1 + \Delta\gamma e^{W_I(t)} - \Delta\phi\right) = & \log(A_2(t)) + \log\left[1 + \left(\frac{1 + \gamma\Delta e^{W_I(t)} - \phi\Delta}{A_2(t)} - 1\right)\right] \\
\approx & \log(\hat{A}_2(t)) + \left(\frac{1 + \gamma\Delta e^{W_I(t)} - \phi\Delta}{\hat{A}_2(t)} - 1\right) \\
& - \frac{1}{2}\left(\frac{1 + \gamma\Delta e^{W_I(t)} - \phi\Delta}{\hat{A}_2(t)} - 1\right)^2.
\end{aligned}
\tag{94}
$$

Then we expand $e^{W_I(t)}$ using a Taylor series to second order:

$$
e^{W_I(t)} \approx e^{\hat{A}_6(t)} + e^{\hat{A}_6(t)}(W_I(t) - \hat{A}_6(t)) + \frac{e^{\hat{A}_6(t)}}{2}(W_I(t) - \hat{A}_6(t))^2.
\tag{95}
$$

Upon substituting (95) into (94), we obtain:

$$
\begin{aligned}
\log\left(1 + \gamma\Delta e^{W_I(t)} - \phi\Delta\right) \approx & \log\hat{A}_2(t) + \frac{1}{\hat{A}_2(t)} + \frac{\gamma\Delta}{\hat{A}_2(t)}e^{\hat{A}_6(t)}(1 - \hat{A}_6(t)) - \frac{\phi\Delta}{\hat{A}_2(t)} - 1 \\
& + \frac{\gamma\Delta e^{\hat{A}_6(t)}}{\hat{A}_2(t)}W_I(t).
\end{aligned}
\tag{96}
$$

Consider $\mu_I^W(t)$ given by (28). Using a Taylor-series expansion to second order, the second term on the right-hand side is:

$$
\log\left(1 - \gamma\Delta + \frac{\beta\Delta\lambda_N e^{W_S(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}}\right)
$$

$$
= \log(A_3(t)) + \log\left[1 + \left(\frac{1 - \gamma\Delta + \frac{\beta\Delta\lambda_N e^{W_S(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}}}{A_3(t)} - 1\right)\right]
$$

$$
\approx \log(\hat{A}_3(t)) + \left(\frac{1 - \gamma\Delta + \frac{\beta\Delta\lambda_N e^{W_S(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}}}{\hat{A}_3(t)} - 1\right) - \frac{1}{2}\left(\frac{1 - \gamma\Delta + \frac{\beta\Delta\lambda_N e^{W_S(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}}}{\hat{A}_3(t)} - 1\right)^2,
$$

$$
(97)
$$

Now, $\frac{e^{W_S(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}}$ can be further expanded using a Taylor series to second order:

$$
\frac{e^{W_S(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}} = \frac{1}{1 - \left(-\frac{1 + e^{W_I(t)}}{e^{W_S(t)}}\right)}
$$

$$
\approx \frac{1}{1 - \hat{A}_{10}(t)} + \frac{\left(-\frac{1 + e^{W_I(t)}}{e^{W_S(t)}} - \hat{A}_{10}(t)\right)}{(1 - \hat{A}_{10}(t))^2} + \frac{\left(-\frac{1 + e^{W_I(t)}}{e^{W_S(t)}} - \hat{A}_{10}(t)\right)^2}{(1 - \hat{A}_{10}(t))^3}.
$$

$$
(98)
$$

Then we expand the remaining nonlinear component in (98), $\frac{1 + e^{W_I(t)}}{e^{W_S(t)}}$, using a Taylor series to second order:

$$
\frac{1 + e^{W_I(t)}}{e^{W_S(t)}} = e^{W_I(t) - W_S(t)} + e^{-W_S(t)}
$$

$$
\approx \left[e^{\hat{A}_4(t)} + e^{\hat{A}_4(t)}(W_I(t) - W_S(t) - \hat{A}_4(t)) + \frac{e^{\hat{A}_4(t)}}{2}(W_I(t) - W_S(t) - \hat{A}_4(t))^2\right]
$$

$$
+ \left[e^{\hat{A}_5(t)} + e^{\hat{A}_5(t)}(-W_S(t) - \hat{A}_5(t)) + \frac{e^{\hat{A}_5(t)}}{2}(-W_S(t) - \hat{A}_5(t))^2\right].
$$

$$
(99)
$$

Then we rewrite (99) as

$$
\frac{1 + e^{W_I(t)}}{e^{W_S(t)}} = B_3(t) + B_4(t)W_S(t) + B_5(t)W_I(t)
$$

$$
+ o\left(W_S(t)^2\right) + o\left(W_I(t)^2\right) + o\left(W_S(t)W_I(t)\right), \qquad (100)
$$

where $\{B_i(t) : i = 3, 4, 5\}$ are defined in (49)-(51).

Hence, combining (97), (98), and (100), we can finally approximate

$$
\log\left[(1 - \gamma\Delta) + \frac{\beta\Delta\lambda_N e^{W_S(t)}}{1 + e^{W_S(t)} + e^{W_I(t)}}\right]
$$

in equation (28) with

$$\log\hat{A}_3(t) + \frac{1 - \gamma\Delta}{\hat{A}_3(t)} + \frac{\beta\Delta\lambda_N}{\hat{A}_3(t)(1 - \hat{A}_{10}(t))} - \frac{\beta\Delta\lambda_N(B_3(t) + \hat{A}_{10}(t))}{\hat{A}_3(t)(1 - \hat{A}_{10}(t))^2} - 1$$

$$+ \left[\frac{-\beta\Delta\lambda_N B_4(t)}{\hat{A}_3(t)(1 - \hat{A}_{10}(t))^2}\right] W_S(t) + \left[1 - \frac{\beta\Delta\lambda_N B_5(t)}{\hat{A}_3(t)(1 - \hat{A}_{10}(t))^2}\right] W_I(t). \quad (101)$$

The third component on the right-hand side of (28) is identical to that of (27), and hence we use the linear approximation given by (96).

Therefore, (93), (96), and (101) yields the linear dynamical process $\boldsymbol{\mu}^{LW}(t)$ as defined in (38)-(39) in Section 3. That is, $\boldsymbol{\mu}^{LW}(t)$ approximates the nonlinear dynamical process, $\boldsymbol{\mu}^{W}(t)$, in the HSIRS model defined by equations (27)-(28).

A.3

Here we give the full conditional distributions that define our Markov chain Monte Carlo (MCMC) algorithm. We use "rest" to represent all other unknowns as well as the data $\{Z_S(1), ..., Z_S(T)\}$ and $\{Z_I(1), ..., Z_I(T)\}$. Notice that this is not a particle-filtering approach in which new data are used to update current and past posteriors without re-running the MCMC.

$- \beta$

$$[\beta|\text{rest}] \propto \prod_{t=2}^{T+F} [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta_\gamma, \phi, \sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2][\beta].$$

$- \theta_\gamma$

$$[\theta_\gamma|\text{rest}] \propto \prod_{t=2}^{T+F} [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta_\gamma, \phi, \sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2][\theta_\gamma].$$

$- \phi$

$$[\phi|\text{rest}] \propto \prod_{t=2}^{T+F} [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta_\gamma, \phi, \sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2][\phi].$$

$- \sigma_{\zeta_S}^2$

$$[\sigma_{\zeta_S}^2|\text{rest}] \propto [W_S(1), W_I(1)|\sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2]$$

$$\cdot \prod_{t=2}^{T+F} [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta_\gamma, \phi, \sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2][\sigma_{\zeta_S}^2].$$

– $\sigma_{\zeta_I}^2$

$$[\sigma_{\zeta_I}^2|\text{rest}] \propto [W_S(1), W_I(1)|\sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2]$$
$$\cdot \prod_{t=2}^{T+F} [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta_\gamma, \phi, \sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2][\sigma_{\zeta_I}^2].$$

– $\{W_S(t), W_I(t)\}$: for $t = 1, ..., T, T+1, ..., T+F$,

$$[W_S(t), W_I(t)|\text{rest}] \propto \begin{cases} [Z_S(t)|W_S(t), W_I(t)] \, [Z_I(t)|W_S(t), W_I(t)] \\ \cdot [W_S(t+1), W_I(t+1)|W_S(t), W_I(t), \beta, \theta_\gamma, \phi, \sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2] \\ \cdot [W_S(t), W_I(t)|\sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2], \\ \quad \text{if } t = 1; \\ [Z_S(t)|W_S(t), W_I(t)] \, [Z_I(t)|W_S(t), W_I(t)] \\ \cdot [W_S(t+1), W_I(t+1)|W_S(t), W_I(t), \beta, \theta_\gamma, \phi, \sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2] \\ \cdot [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta_\gamma, \phi, \sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2], \\ \quad \text{if } 2 \le t \le T; \\ \left[W_S(t+1), W_I(t+1)|W_S(t), W_I(t), \beta, \theta_\gamma, \phi, \sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2 \right] \\ \cdot \left[W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta_\gamma, \phi, \sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2 \right], \\ \quad \text{if } T < t < T+F; \\ \cdot [W_S(t), W_I(t)|W_S(t-1), W_I(t-1), \beta, \theta_\gamma, \phi, \sigma_{\zeta_S}^2, \sigma_{\zeta_I}^2], \\ \quad \text{if } t = T+F. \end{cases}$$

References

Allen, L. (2003). *An Introduction to Stochastic Processes with Applications to Biology.* Prentice-Hall, Upper Saddle River, NJ.

Alonso, D., McKane, A. J., and Pascual, M. (2007). Stochastic amplification in epidemics. *Journal of the Royal Society Interface*, 4(14):575–582.

Anderson, R. (2006). Planning for pandemics of infectious diseases. *The Bridge*, 3:5–9.

Anderson, R. and May, R. (1991). *Infectious Diseases of Humans: Dynamics and Control.* Oxford University Press, Oxford, UK.

Andersson, H. and Britton, T. (2000). *Stochastic Epidemic Models and Their Statistical Analysis*, volume 151. Springer, Lecture Notes in Statistics.

Arulampalam, S., Maskell, S., Gordon, N., and Clapp, T. (2001). A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50:174–188.

Bailey, N. T. J. (1975). *The Mathematical Theory of Infectious Diseases.* Griffin, London.

Bansal, S., Pourbohloul, B., Hupert, N., Grenfell, B., and Meyers, L., A. (2010). The shifting demographic landscape of pandemic influenza. *PLoS ONE*, 5(2):e9360.doi:10.1371/hournal.phone.0009360.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43:1–59.

Burr, T. and Chowell, G. (2006). Observation and model error effects on parameter estimates in susceptible-infected-recovered epidemiological models. *Far East Journal of Theoretical Statistics*, 19:163–183.

Carlin, B. and Banerjee, S. (2002). Hierarchical multivariate CAR models for spatio-temporally correlated survival data (with discussion). In Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Heckerman, D., Smith, A. F. M., and West, M., editors, *in Bayesian Statistics 7*, pages 45–63. Clarendon Press, Oxford, UK.

Catterall, S., Cook, A. R., Marion, G., Butler, A., and Hulme, P. E. (2012). Accounting for uncertainty in colonisation times: a novel approach to modelling the spatio-temporal dynamics of alien invasions using distribution data. *Ecography*, 35(10):901–911.

Chao, D. L., Bloom, J. D., Kochin, B. F., Antia, R., and Longini, Jr, I. M. (2012). The global spread of drug-resistant influenza. *Journal of the Royal Society Interface*, 9(69):648–656.

Cook, A. R., Otten, W., Marion, G., Gibson, G. J., and Gilligan, C. A. (2007). Estimation of multiple transmission rates for epidemics in heterogeneous populations. *Proceedings of the National Academy of Sciences of the United States of America*, 104(51):20392–20397.

Cressie, N., Shi, T., and Kang, E. (2010). Fixed rank filtering for spatio-temporal data. *Journal of Computational and Graphical Statistics*, 19:724–745.

Cressie, N. and Wikle, C. (2011). *Statistics for Spatio-Temporal Data*. Wiley, Hoboken, NJ.

Dukić, V., Lopes, H. F., and Polson, N. G. (2012). Tracking epidemics with Google flu trends data and a state-space SEIR model. *Journal of the American Statistical Association*, 107:1410–1426.

Dushoff, J., Plotkin, J., Levin, S., and Earn, D. (2004). Dynamical resonance can account for seasonality of influenza epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 101:16915–16916.

Fine, P., Eames, K., and Heymann, D. L. (2011). "Herd immunity": A rough guide. *Clinical Infectious Diseases*, 52(7):911–916.

Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica*, 6:733–807.

Gibson, G. J., Otten, W., Filipe, J. A. N., Cook, A., Marion, G., and Gilligan, C. A. (2006). Bayesian estimation for percolation models of disease spread in plant populations. *Statistics and Computing*, 16(4):391–402.

Halloran, M. E., Longini, I. M., Nizam, A., and Yang, Y. (2002). Containing bioterrorist smallpox. *Science*, 298:1428–1432.

He, D., Ionides, E. L., and King, A. A. (2010). Plug-and-play inference for disease dynamics: measles in large and small populations as a case study.

Journal of the Royal Society Interface, 7:271–283.

Hethcote, H. (2000). The mathematics of infectious diseases. *SIAM Review*, 42:599–653.

Hooten, M. B. and Wikle, C. K. (2010). Statistical agent-based models for discrete spatio-temporal systems. *Journal of the American Statistical Association*, 105:236–248.

Ionides, E. L., Bretó, C., and King, A. A. (2006). Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences of the United States of America*, 103:18438–18443.

Keeling, M. and Rohani, P. (2007). *Modelling Infectious Diseases in Humans and Animals.* Princeton University Press, Princeton, NJ.

Kendall, B. E., Briggs, C. J., Murdoch, W. W., Turchin, P., Ellner, S. P., McCauley, E., Nisbet, R. M., and Wood, S. N. (1999). Why do populations cycle? a synthesis of statistical and mechanistic modeling approaches. *Ecology*, 80:1789–1805.

Kendall, B. E., Ellner, S. P., McCauley, E., Wood, S. N., Briggs, C. J., Murdoch, W. M., and Turchin, P. (2005). Population cycles in the pine looper moth: dynamical tests of mechanistic hypotheses. *Ecological Monographs*, 75:259–276.

Kermack, W. and McKendrick, A. (1927). Contributions to the mathematical theory of epidemics - I. *Proceedings of the Royal Society of Edinburgh A*, 115:700–721.

Kermack, W. and McKendrick, A. (1932). Contributions to the mathematical theory of epidemics - II-the problem of endemicity. *Proceedings of the Royal Society of Edinburgh A*, 138:55–83.

Kermack, W. and McKendrick, A. (1933). Contributions to the mathematical theory of epidemics - III-further studies of the problem of endemicity. *Proceedings of the Royal Society of Edinburgh A*, 141:94–122.

King, A. A., Ionides, E. L., Pascual, M., and Bouma, M. J. (2008). Inapparent infections and cholera dynamics. *Nature*, 454:877–880.

Liu, J. and West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In Doucet, A., de Freitas, N., and Gordon, N. J., editors, *Sequential Monte Carlo Methods in Practice*, pages 197–224. Springer, New York.

Lloyd, A. (2001). Realistic distributions of infectious periods in epidemic models: Changing patterns of persistence and dynamics. *Theoretical Population Biology*, 60:59–71.

Mugglin, A., Cressie, N., and Gemmell, I. (2002). Hierarchical statistical modelling of influenza epidemic dynamics in space and time. *Statistics in Medicine*, 21:2703–2721.

Oleson, J. J. and Wikle, C. K. (2013). Predicting infectious disease outbreak risk via migratory waterfowl vectors. *Journal of Applied Statistics*, 40(3):656–673.

Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., and Matthiopoulos, J. (2008). State-space models of individual animal movement. *Trends in Ecology & Evolution*, 23(2):87–94.

Potter, C. W. (2001). A history of influenza. *Journal of Applied Microbiology*, 91:572–579.

Rasmussen, D. A., Ratmann, O., and Koelle, K. (2011). Inference for nonlinear epidemiological models using genealogies and time series. *PLoS Computational Biology*, 7(8):e1002136.

Reluga, T. (2004). A two-phase epidemic driven by diffusion. *Journal of Theoretical Biology*, 229:249–261.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods, second edition.* Springer, New York, NY.

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

Volz, E. (2008). SIR dynamics in random networks with heterogeneous connectivity. *Journal of Mathematical Biology*, 56:293–310.

Waller, L., Carlin, B., Xia, H., and Gelfand, A. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, 92:607–617.

Wearing, H., Rohani, P., and Keeling, M. (2005). Appropriate models for the management of infectious diseases. *PLoS Medicine*, 2(7):e174.

Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466:1102–1104.

Xu, Y., Allena, L., and Perelson, A. (2007). Stochastic model of an influenza epidemic with drug resistance. *Journal of Theoretical Biology*, 248:179–193.

Zhou, T., Fu, Z., and Wang, B. (2006). Epidemic dynamics on complex networks. *Progress in Natural Science*, 16:452–457.

Zhuang, L., Cressie, N., Pomeroy, L., and Janies, D. (2013). Multi-species SIR models from a dynamical Bayesian perspective. *Theoretical Ecology*, 6(4):457–473.

Tables

Table 1 Table of the initializations $\{\hat{A}_i(t) : i = 1, ..., 10\}$ and the quantities $\{A_i(t) : i = 1, ..., 10\}$ that they approximate.

Value	Initializations ($\hat{A}(t)$)	Nonlinear Components in HSIRS
$A_1(t)$	$1 + \frac{\Delta\phi_0 Z_R(t)}{Z_S(t)} - \Delta\beta_0 Z_I(t)$	$1 + \frac{\Delta\phi}{e^{W_S(t)}} - \frac{\Delta\beta\lambda_N e^{W_I(t)}}{1+e^{W_S(t)}+e^{W_I(t)}}$
$A_2(t)$	$1 + \Delta\gamma_0 \frac{Z_I(t)}{Z_R(t)} - \Delta\phi_0$	$1 + \Delta\gamma e^{W_I(t)} - \Delta\phi$
$A_3(t)$	$(1 - \Delta\gamma_0) + \Delta\beta_0 Z_S(t)$	$(1 - \Delta\gamma) + \frac{\Delta\beta\lambda_N e^{W_S(t)}}{1+e^{W_S(t)}+e^{W_I(t)}}$
$A_4(t)$	$\log\frac{Z_I(t)}{Z_S(t)}$	$(W_I(t) - W_S(t))$
$A_5(t)$	$-\log\frac{Z_S(t)}{Z_R(t)}$	$-W_S(t)$
$A_6(t)$	$\log\frac{Z_I(t)}{Z_R(t)}$	$W_I(t)$
$A_7(t)$	$-\frac{Z_I(t)}{(Z_N - Z_I(t))}$	$-\frac{e^{W_I(t)}}{(1+e^{W_S(t)})}$
$A_8(t)$	$-\frac{Z_S(t)}{Z_R(t)}$	$-e^{W_S(t)}$
$A_9(t)$	$-\frac{Z_R(t)}{Z_S(t)}$	$-e^{-W_S(t)}$
$A_{10}(t)$	$-\frac{Z_N - Z_S(t)}{Z_S(t)}$	$\frac{-(1+e^{W_I(t)})}{e^{W_S(t)}}$

Table 2 Table of parameters and summaries of their priors and posteriors for the case AH.

Parameter	Prior distribution	Prior quantiles			Prior	Posterior quantiles				Posterior
		0.025	0.5	0.975	standard deviation	0.025	0.5	0.975	mean	standard deviation
$\beta \times 10^6$	Uniform$[0, \beta_{max} \times 10^6 = 6.7686]$	0.1692	3.384	6.599	1.9545	0.2244	0.4763	0.7260	0.4769	0.1257
θ_γ	Normal(-0.708 ,0.01)	-0.904	-0.708	-0.512	0.1001	-0.897	-0.707	-0.515	-0.707	0.0096
ϕ	Uniform[0,1]	0.025	0.5	0.975	0.2889	0.0071	0.0836	0.1996	0.0876	0.0026
$\sigma^2_{\xi_S}$	Inverse Gamma(0.25, 0.4)	1.457	57.2	9.48e6	2.84e24	0.1546	0.2599	0.4878	0.2758	0.0075
$\sigma^2_{\xi_I}$	Inverse Gamma(0.25, 0.4)	1.457	57.2	9.48e6	2.84e24	0.1357	0.2276	0.4198	0.2405	0.0055

Table 3 Table of parameters and summaries of their priors and posteriors for the case AM.

Parameter	Prior distribution	Prior quantiles			Prior	Posterior quantiles				Posterior
		0.025	0.5	0.975	standard deviation	0.025	0.5	0.975	mean	standard deviation
$\beta \times 10^6$	Uniform$[0, \beta_{max} \times 10^6 = 9.7523]$	0.2438	4.876	9.508	2.816	0.2471	0.4814	0.7241	0.4820	0.1209
θ_γ	Normal(-0.708 ,0.01)	-0.904	-0.708	-0.512	0.1001	-0.9009	-0.7009	-0.5095	-0.7023	0.0099
ϕ	Uniform[0,1]	0.025	0.5	0.975	0.2889	0.0060	0.0707	0.1711	0.0749	0.0020
$\sigma^2_{\xi_S}$	Inverse Gamma(0.25, 0.4)	1.457	57.2	9.48e6	2.84e24	0.1246	0.2074	0.3850	0.2194	0.0046
$\sigma^2_{\xi_I}$	Inverse Gamma(0.25, 0.4)	1.457	57.2	9.48e6	2.84e24	0.1248	0.2099	0.3888	0.2217	0.0047

Table 4 Analysis of variance (ANOVA) on $A_S^C(T^O) - A_S^H(T^O)$ up to two-way interactions. In the table, DF denotes degrees of freedom; SS denotes sums of squares; MS denotes mean squares; and F value denotes the value of the F statistic. Relatively large F values are highlighted in gray.

Source	DF	SS	MS	F value
DI	1	0.014403	0.014403	9040.97
SV	1	0.04542	0.04542	28511.79
TR	1	0.005874	0.005874	3686.90
DI× SV	1	0.001107	0.001107	694.92
DI× TR	1	0.001098	0.001098	689.03
SV× TR	1	0.000171	0.000171	107.24
Residuals	1	0.000002	0.000002	
Total	7	0.06808		

Table 5 Analysis of variance (ANOVA) on $A_I^C(T^O) - A_I^H(T^O)$ up to two-way interactions. In the table, DF denotes degrees of freedom; SS denotes sums of squares; MS denotes mean squares; and F value denotes the value of the F statistic. Relatively large F values are highlighted in gray.

Source	DF	SS	MS	F value
DI	1	0.0013685	0.0013685	45.2129
SV	1	0.0076848	0.0076848	253.8924
TR	1	0.0071008	0.0071008	234.5970
DI× SV	1	0.0000914	0.0000914	3.0193
DI× TR	1	0.0001487	0.0001487	4.9117
SV× TR	1	0.0001683	0.0001683	5.5596
Residuals	1	0.0000303	0.0000303	
Total	7	0.01659275		

Table 6 Analysis of variance (ANOVA) on $A_S^C(T^M) - A_S^H(T^M)$ up to two-way interactions. In the table, DF denotes degrees of freedom; SS denotes sums of squares; MS denotes mean squares; and F value denotes the value of the F statistic. Relatively large F values are highlighted in gray.

Source	DF	SS	MS	F value
DI	1	0.0040394	0.0040394	7.18
SV	1	0.0008263	0.0008263	1.47
TR	1	0.0001459	0.0001459	0.26
DI× SV	1	0.0000168	0.0000168	0.03
DI× TR	1	0.0001824	0.0001824	0.32
SV× TR	1	0.0008041	0.0008041	1.43
Residuals	1	0.0005623	0.0005623	
Total	7	0.006577339		

Table 7 Analysis of variance (ANOVA) on $A_I^C(T^M) - A_I^H(T^M)$ up to two-way interactions. In the table, DF denotes degrees of freedom; SS denotes sums of squares; MS denotes mean squares; and F value denotes the value of the F statistic. Relatively large F values are highlighted in gray.

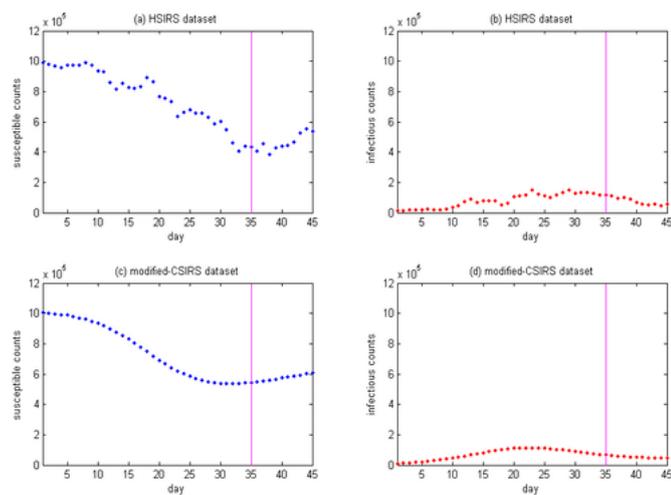| Source	DF	SS	MS	F value
DI	1	0.00001648	0.00001648	0.10
SV	1	0.00011556	0.00011556	0.71
TR	1	0.00021602	0.00021602	1.33
DI× SV	1	0.00016583	0.00016583	1.02
DI× TR	1	0.00005301	0.00005301	0.33
SV× TR	1	0.00067383	0.00067383	4.15
Residuals	1	0.00016244	0.00016244	
Total	7	0.001403168		

Figures

Fig. 1 Plots of counts $\{\mathbf{Z}(t)\}$ for the HSIRS and the modified-CSIRS datasets as a function of time. The left-hand plots (a) and (c) show $\{Z_s(t)\}$, and the right-hand plots (b) and (d) show $\{Z_I(t)\}$. The vertical line in each plot emphasizes that for the analysis the data are only available on days 1-35, and they are missing on days 36-45.
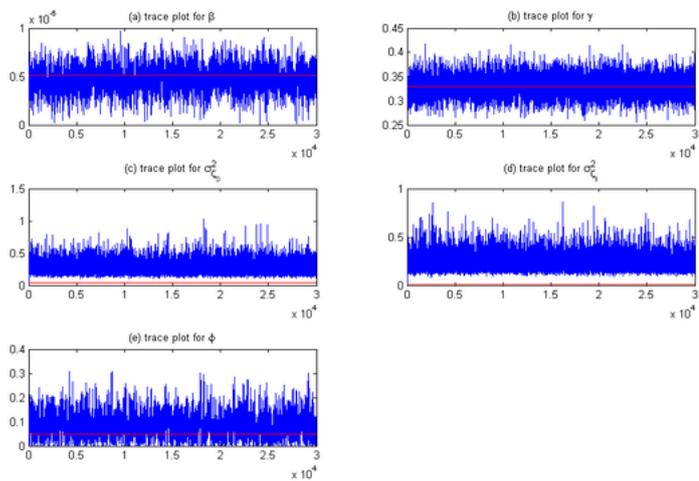
Fig. 2 Trace plots for the case AH: (a) trace plot for β; (b) trace plot for γ; (c) trace plot for $\sigma^2_{\zeta_S}$; (d) trace plot for $\sigma^2_{\zeta_I}$; (e) trace plot for ϕ. Red lines in plots (a)-(e) indicate the true values of the parameters used for simulating the HSIRS dataset, namely, β^*, γ^*, $\sigma^{2*}_{\xi_S}$, $\sigma^{2*}_{\xi_I}$, and ϕ^*, respectively
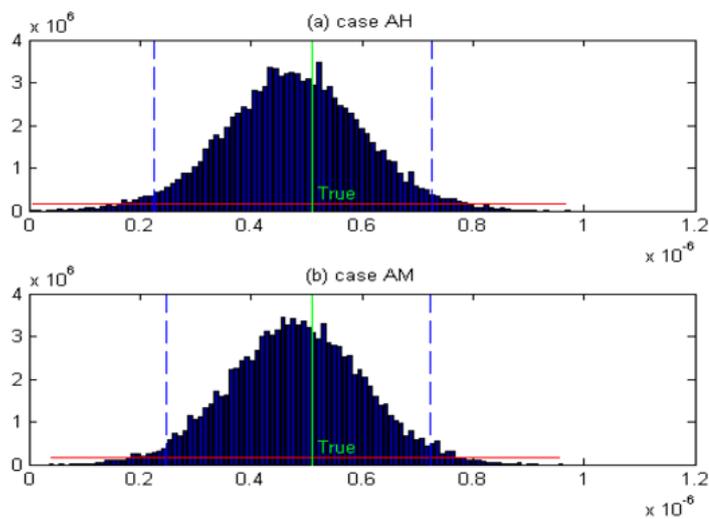
Fig. 3 Prior (red line) and Posterior (histogram) distributions of β for (a) case AH; (b) case AM. In each plot, blue dashed lines indicate 95% posterior credible intervals; the green solid line indicates the true value.
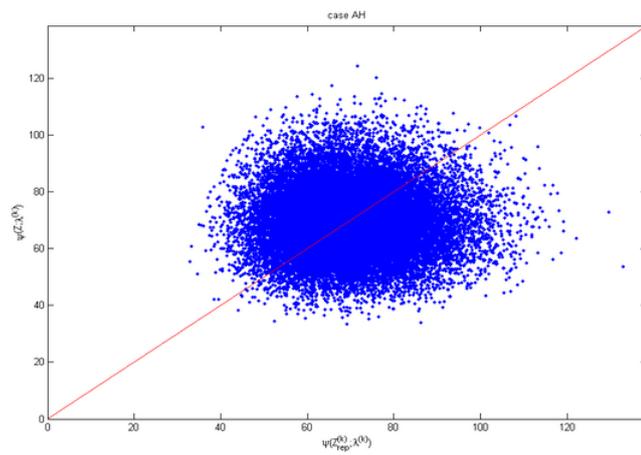
Fig. 4 Scatter plot of $\psi\left(\mathbf{Z}; \boldsymbol{\lambda}^{(k)}\right)$ versus $\psi\left(\mathbf{Z}_{rep}^{(k)}; \boldsymbol{\lambda}^{(k)}\right)$ for the case AH.
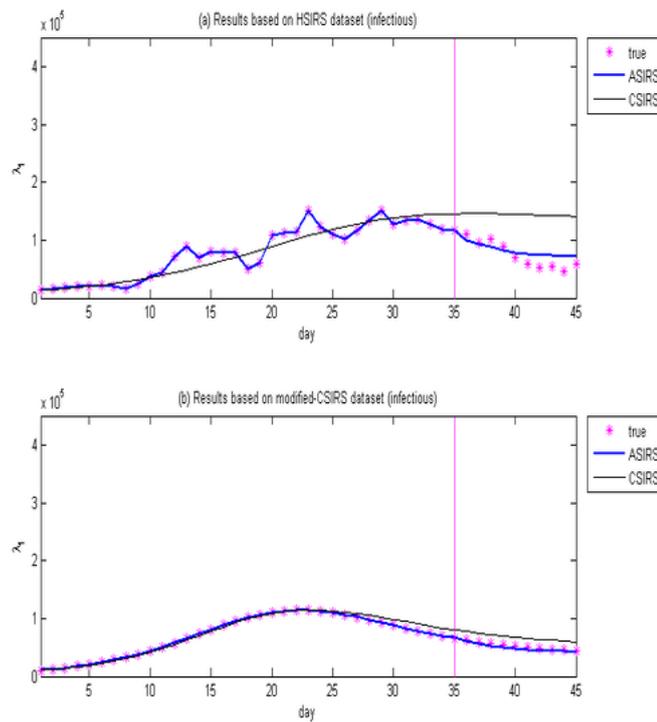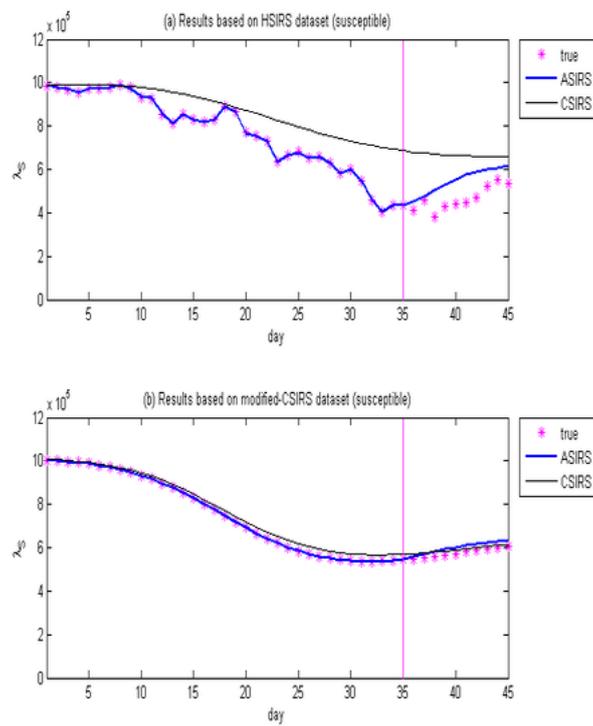
Fig. 5 Estimated infectious counts obtained from fitting the ASIRS model and the CSIRS model to the HSIRS and modified-CSIRS datasets, as a function of time. The vertical line in each plot emphasizes that data are only available on days 1-35, and that they are missing on days 36-45. The blue solid line is the posterior median of $\{\lambda_I(t)\}$ obtained from the ASIRS model. The black thin line is the estimate of the infectious counts from the CSIRS model. The pink stars give the true infectious counts.
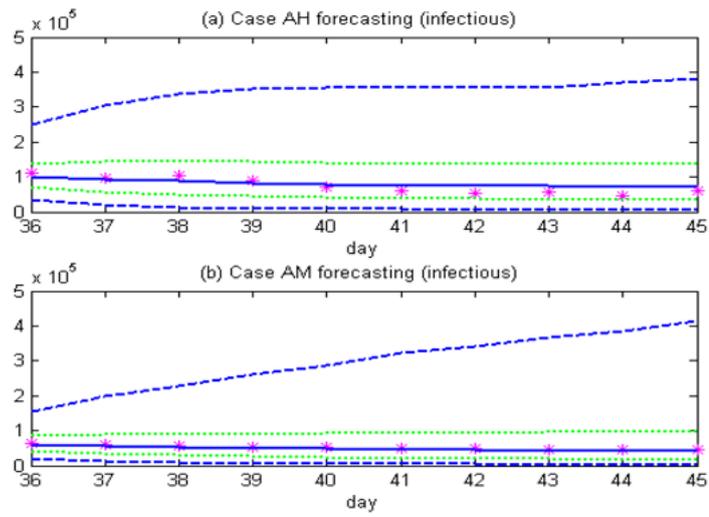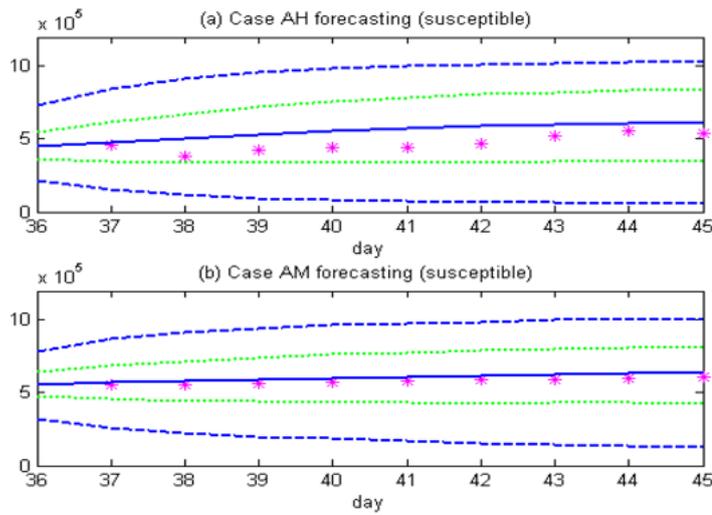
Fig. 6 Posterior quantiles of $\{\lambda_I(t)\}$ for the infectious population, obtained from the ASIRS model during the forecasting period from day 36 to day 45 (based on data from day 1 to day 35, from the respective datasets). Lines (from bottom to top) indicate the 0.025, 0.25, 0.5, 0.75 and 0.975 quantiles, and pink stars give the true values of $\{\lambda_I(t)\}$.

Fig. 7 Estimated susceptible counts obtained from fitting the ASIRS model and the CSIRS model to the HSIRS and modified-CSIRS datasets, as a function of time. The vertical line in each plot emphasizes that data are only available on days 1-35, and that they are missing on days 36-45. The blue solid line is the posterior median of $\{\lambda_S(t)\}$ obtained from the ASIRS model. The black thin line is the estimate of the infectious counts from the CSIRS model. The pink stars give the true susceptible counts.

Fig. 8 Posterior quantiles of $\{\lambda_S(t)\}$ for the susceptible population, obtained from the ASIRS model during the forecasting period from day 36 to day 45 (based on data from day 1 to day 35, from the respective datasets). Lines (from bottom to top) indicate the 0.025, 0.25, 0.5, 0.75 and 0.975 quantiles, and pink stars give the true values of $\{\lambda_S(t)\}$.
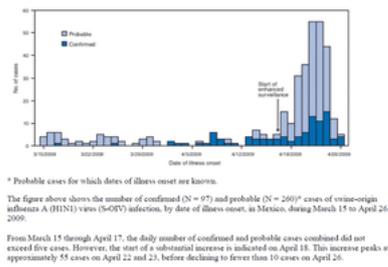
Fig. 9 A figure from "Outbreak of Swine-Origin Influenza A (H1N1) Virus Infection—Mexico, March–April 2009," *Centers for Disease Control and Prevention (CDC)*.
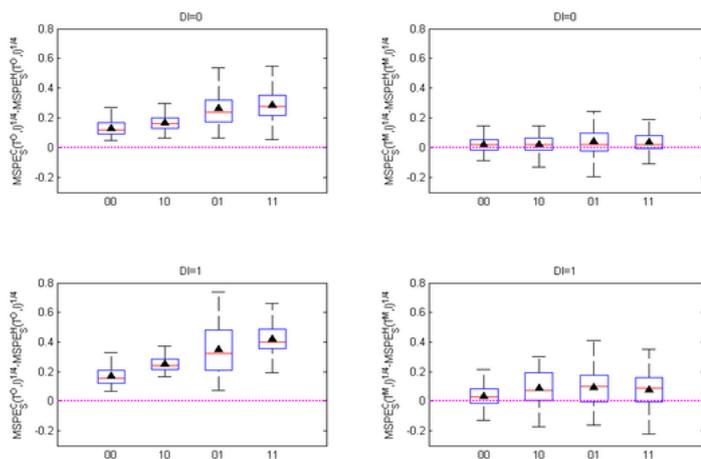
Fig. 10 Boxplots of differences, $\left\{MSPE_S^C(T^*,l)^{1/4} - MSPE_S^H(T^*,l)^{1/4} : l = 1, ..., L\right\}$, for the susceptible population, at given levels of **DI** (Data Information), **TR** (Transmission Rate), and **SV** (Small-scale Variation). Left panels: $T^* = T^O$; right panels: $T^* = T^M$. Upper panels: DI=0; lower panels: DI=1. The x-axis labels are 'ab', where 'a' (=0 and 1) represents the level for the factor **TR**, and 'b' (=0 and 1) represents the level for the factor **SV**. The triangles show the mean of the differences.
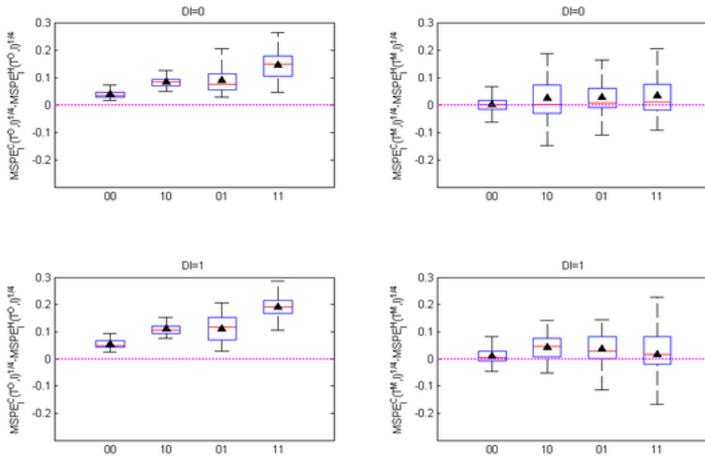
Fig. 11 Boxplots of differences, $\left\{MSPE_I^C(T^*,l)^{1/4} - MSPE_I^H(T^*,l)^{1/4} : l = 1, ..., L\right\}$, for the infectious population, at given levels of **DI** (Data Information), **TR** (Transmission Rate), and **SV** (Small-scale Variation). Left panels: $T^* = T^O$; right panels: $T^* = T^M$. Upper panels: given DI=0; lower panels: given DI=1. The x-axis label 'ab', where 'a' (=0 and 1) represents the level for the factor **TR**; and 'b' (=0 and 1) represents the level for the factor **SV**. The triangles show the mean of the differences.
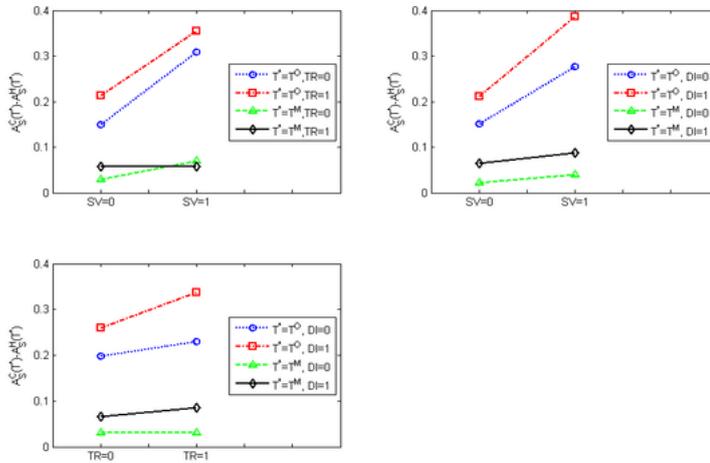
Fig. 12 Plots showing interaction for the susceptible population, between **SV** (Small-scale Variation) and **TR** (Transmission Rate), between **SV** and **DI** (Data Information), and between **TR** and **DI** in the ANOVA of $A_S^C(T^*) - A_S^H(T^*)$ (red and blue lines: $T^* = T^O$; green and black lines: $T^* = T^M$).
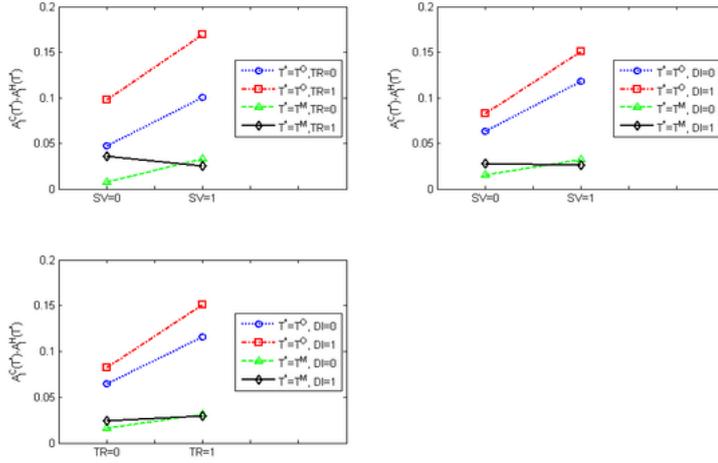
Fig. 13 Plots showing interaction for the infectious population between **SV** (Small-scale Variation) and **TR** (Transmission rate), between **SV** and **DI** (Data Information), and between **TR** and **DI** in the ANOVA of $A_I^C(T^*) - A_I^H(T^*)$ (red and blue lines: $T^* = T^O$; green and black lines: $T^* = T^M$).
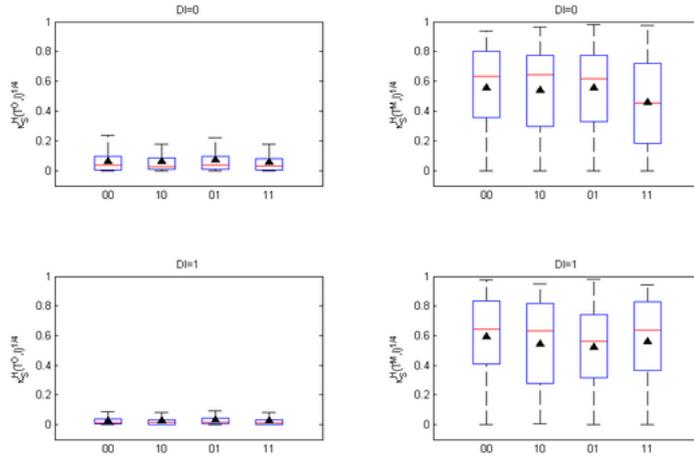

Fig. 14 Boxplots of fractions of MSPE explained by squared bias, $\left\{ \kappa_S^H(T^*, l) : l = 1, ..., L \right\}$, for the susceptible population, at given levels of **DI** (Data Information), **TR** (Transmission Rate), and **SV** (Small-scale Variation). Left panels: $T^* = T^O$; right panels: $T^* = T^M$. Upper panels: DI=0; lower panels: DI=1. The x-axis labels are 'ab', where 'a' (=0 and 1) represents the level for the factor **TR**, and 'b' (=0 and 1) represents the level for the factor **SV**. The triangles show the mean of the fraction.
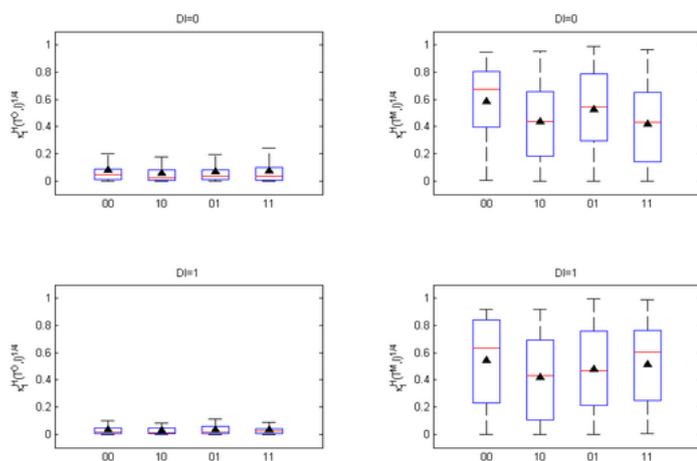
Fig. 15 Boxplots of fractions of MSPE explained by squared bias, $\left\{\kappa_I^H(T^*,l) : l = 1, ..., L\right\}$, for the infectious population, at given levels of **DI** (Data Information), **TR** (Transmission Rate), and **SV** (Small-scale Variation). Left panels: $T^* = T^O$; right panels: $T^* = T^M$. Upper panels: given DI=0; lower panels: given DI=1. The x-axis label 'ab', where 'a' (=0 and 1) represents the level for the factor **TR**; and 'b' (=0 and 1) represents the level for the factor **SV**. The triangles show the mean of the fractions.