The Canadian Journal of Statistics
Vol. 44, No. 4, 2016, Pages 480–501
La revue canadienne de statistique

Gaussian process emulators for spatial individual-level models of infectious disease

Gyanendra POKHAREL[1,2]* and Rob DEARDON[1,3]*

[1]*Department of Mathematics and Statistics, Faculty of Science, University of Calgary, Alberta, Canada*
[2]*Department of Mathematics and Statistics, College of Physical and Engineering Science, University of Guelph, Ontario, Canada*
[3]*Department of Production Animal Health, Faculty of Veterinary Medicine, University of Calgary, Alberta, Canada*

Key words and phrases: Disease models; emulators; Gaussian process; spatial models; tomato spotted wilt virus.

MSC 2010: Primary 92B15; secondary 62F15

Abstract: Statistical inference for spatial models of infectious disease spread is often very computationally expensive. These models are generally fitted in a Bayesian Markov chain Monte Carlo (MCMC) framework, which requires multiple iterations of the computationally cumbersome likelihood function. We here propose a method of inference based on so-called emulation techniques. Once again the method is set in a Bayesian MCMC context, but avoids calculation of the computationally expensive likelihood function by replacing it with a Gaussian process approximation of the likelihood function built from simulated data. We show that such a method can be used to infer the model parameters and underlying characteristics of the spatial disease system, and this can be done in a computationally efficient manner. *The Canadian Journal of Statistics* 44: 480–501; 2016 © 2016 Statistical Society of Canada

Résumé: L'inférence statistique avec des modèles spatiaux pour la transmission de maladies infectieuses exige souvent de lourds calculs informatiques. Ces modèles s'inscrivent généralement dans un cadre de méthodes de Monte Carlo par chaînes de Markov (MCMC) qui nécessite plusieurs répétitions d'un long calcul de vraisemblance complexe. Les auteurs proposent une méthode d'inférence basée sur des techniques d'émulation. La méthode demeure dans un cadre de MCMC bayésien, mais les onéreux calculs de vraisemblance sont évités en la remplaçant par une approximation basée sur des processus gaussiens ajustés sur des données simulées. Les auteurs montrent que leur méthode basée sur des calculs rapides peut servir à inférer les paramètres d'un système sous-jacent de transmission spatiale de la maladie. *La revue canadienne de statistique* 44: 480–501; 2016 © 2016 Société statistique du Canada

1. INTRODUCTION

Epidemiologists have a great deal of interest in the mathematical modelling of infectious disease spread dynamics, as resulting models can help us to understand underlying disease transmission mechanisms. Understanding such mechanisms might be crucial for the emergency management of disease, helping to develop disease control strategies via a fitted model. Further in many infectious disease systems, spatial distance between individuals is a key risk factor to account for in the model (e.g., Hughes et al., 1997; Chis-Ster & Ferguson, 2007; Deardon et al., 2010; Kwong & Deardon, 2012; Jandarov et al., 2014; Pokharel & Deardon, 2014).

Deardon et al. (2010) proposed a class of discrete-time individual level models (ILMs) that can be fitted to space-time infectious disease data. These models can aid greatly in increasing

* *Author to whom correspondence may be addressed.*
E-mail: gyanendra.pokharel@ucalgary.ca, robert.deardon@ucalgary.ca

understanding of complex disease systems. However, parameter estimation for spatial ILMs can be computationally burdensome. In particular, repeated calculation of the likelihood function in a Markov chain Monte Carlo (MCMC) algorithm (e.g., Gamerman & Lopes, 2006)—as is typically used—can result in a computationally intractable analysis for large data. This poses especially acute problems when the goal of fitting an infectious disease model is to analyze data collected from an emerging epidemic with the hope of using the model to guide control and mitigation strategies (e.g., Haydon et al., 2003; Tildesley et al., 2006; Cook et al., 2008).

In such cases it is necessary to resort to some sort of alternative analysis, accurately informative enough to afford practically useful information in a reasonable time frame. Often such inference is based on a computationally tractable approximation of the posterior rather than the true posterior. One avenue of research of great current interest is in the use of so-called approximate Bayesian computation (ABC) methods (e.g., McKinley, Cook, & Deardon, 2009; Numminen et al., 2013). In such methods the likelihood is approximated by comparing simulated data to real data, usually via some summary metrics, within the framework of, say, an MCMC (e.g., Marjoram et al., 2003) or sequential Monte Carlo (e.g., Beaumont et al., 2009) algorithm. A number of other methods have been developed in this regard; for example, methods based upon auxiliary variables (O'Neill & Roberts, 1999), particle filtering (Arulampalam et al., 2002), and partially observed Markov process (POMP) (King, Nguyen, & Ionides, 2015).

We here consider approximations based upon the idea of the "emulation" of simulation models (e.g., O'Hagan & Kingman, 1978; Sacks et al., 1989; Currin et al., 1991; Craig et al., 2001; Kennedy & O'Hagan, 2001; Bayarri et al., 2007). An emulator is a stochastic process that represents a simulator, a mathematical representation of a physical system implemented in a computer (Bastos & O'Hagan, 2009). Such emulators are typically used in situations where carrying out the required number of realizations from the simulator (here the ILM likelihood) is computationally intractable.

Given its flexible framework for modelling complex relationships, Gaussian process (GP) emulators are widely used, including for modelling spatio-temporal data (Liu & West, 2004). The popularity of GP stems from two essential properties. First a GP is completely determined by its mean and covariance functions. Second construction of the predictive distribution—the distribution of new observations conditional upon the fitted GP—is relatively straightforward (Davis, 2001). We here consider emulation-based inference in which the goal is to use the predictive distribution of the fitted GP to approximate the mapping between the ILM parameters and the likelihood function.

We adapt a widely used two-stage approach developed for emulators (e.g., Sacks et al., 1989; Kennedy & O'Hagan, 2001; Oakley & O'Hagan, 2002). In the first stage, a GP emulator is built using the log L_2-norm distance between a set of summary statistics of the observed data and simulated data as the response variable of the GP, with the parameter values of the true model (here the ILM) providing the explanatory variables. This is done over a pre-defined design matrix of ILM parameter values (i.e., a set of parameter values for which epidemics are simulated and compared with the observed data). Here, following Jandarov et al. (2014), the summary statistics used are based on the epidemic curve—the number of infections that occur at each of a series of discrete time points. Then in the second stage an approximate Bayesian inference is carried out using MCMC with the predictive distribution of the GP from the first stage used to provide an approximation to the likelihood function of the ILM. As is common we also incorporate an additional parameter allowing for "model-data discrepancy" between the ILM and GP (e.g., Bayarri et al., 2007; Bhat, Haran, & Goes, 2010).

This article has several objectives. The first is to see how well emulation-based methods can infer spatial disease dynamics for epidemic curve data collected from a global population when using ILM-type models of Deardon et al. (2010) as a proof-of-concept. The second is to

compare emulation-based inference based on global epidemic curve summary statistics with that based upon spatially stratified epidemic curves. Spatial stratification here refers to splitting the global population of interest into smaller groups based on the spatial location of individuals, and considering epidemic curve summary statistics for each spatial group. The third objective is to determine the robustness to the resolution of the design matrix used to build the GP emulator. We finally want to observe how computationally efficient these emulation-based methods are compared to the traditional non-approximated Bayesian MCMC method.

The remainder of this article is laid out as follows. Our infectious disease transmission model framework, methods for the construction of the GP emulator, various schemes of spatial stratification, and methods used for model fitting are outlined in Section 2. Section 3 details data used to test our methods, including simulated data and data from an experiment on tomato spotted wilt virus (TSWV). Results from our analyses are presented in Section 4 and finally discussion of conclusions and possible further work are presented in Section 5.

2. MODEL FRAMEWORK

The discrete time-ILM framework for modelling epidemic transmission over time and space proposed by Deardon et al. (2010) is used here. A susceptible–infectious–removed (SIR) version of this framework will be described briefly.

In an SIR framework, an individual, i, can be any one of three states—susceptible (S), infectious (I), or removed (R)—at any discrete time point t. Individuals in the S state are not infected but can contract the disease. Individuals in the I state are infectious and can transmit the disease to S individuals. Individuals in the R state are assumed to have been removed from the S population, for example, through recovery with acquired immunity following infection, death, or quarantine. An individual contracting the disease goes through all three compartments, $S \rightarrow I \rightarrow R$, in order. The number of time points spent in the I state before moving from $I \rightarrow R$ is called the I period, here denoted γ_I.

Let $S(t)$, $I(t)$, and $R(t)$ denote, respectively, the sets of S, I, and R individuals at time t. An individual i must be in one, and only one, of these sets at any time point t. The complete epidemic history consists of those three sets for $t = 1, 2, \ldots, t_{max}$, where $t = 1$ is the time when the first infection is observed and t_{max} is the time when the last infective enters the R state (i.e., the epidemic ends).

2.1. Spatial ILM

In this article we restrict our discussion to a simple spatial ILM in which Euclidean distance between individuals is the only individual-level covariate driving the disease transmission. A more general form of ILMs is given in Deardon et al. (2010), which allows for various other individual level covariates to be included (e.g., contact networks, vaccination status, genetics, lifestyle factors, etc.).

In the model considered here, P_{it}, the probability that a previously uninfected (i.e., S) individual i becomes infected at time t is given by:

$$P_{it} = 1 - \exp\left[-\left\{\alpha \sum_{j \in I(t)} d_{ij}^{-\beta}\right\}\right];\ \alpha, \beta > 0, \tag{1}$$

where α is an infectivity parameter reflecting the overall strength of the epidemic; β is the power-law decay or spatial parameter; and $d_{ij}^{-\beta}$ is a distance-based power-law "infection kernel." The spatial parameter determines the risk of disease transmission over varying distance. We assume a fixed and known I period for all individuals, γ_I, and assume that there are no infections unexplained

by the underlying model (e.g., there are no infections entering the observed population from outside). Note that these two simplifying assumptions can easily be relaxed.

Given the complete epidemic history the likelihood function for model (1) with epidemic data $\mathscr{D}$ and model parameter vector $\theta = (\alpha, \beta)$ is given by

$$f(\mathscr{D}|\theta) = \prod_{t=1}^{t_{max}} \left[\prod_{i \in I(t+1) \setminus I(t)} P_{it} \right] \left[\prod_{i \in S(t+1)} (1 - P_{it}) \right], \tag{2}$$

where $I(t+1)\setminus I(t)$ is the set of newly infected individuals at time point $(t+1)$ and $S(t+1)$ is the set of S individuals at time point $(t+1)$. In situations where we have very large data sets, calculation of this likelihood function can prove computationally intractable (e.g., Deardon et al., 2010).

2.2. GP Emulator

An emulator is a statistical approximation that mimics a larger complex computer model (or simulator [Conti & O'Hagan, 2010]). To be useful evaluations of the emulator should be much faster than evaluations of the complex computer model it approximates. GPs offer a flexible statistical model commonly used as the emulator.

Here it is the calculation of the likelihood function and, thus, the posterior distribution in a Bayesian framework, that can cause the computational bottleneck (e.g., when carrying out MCMC). We therefore propose to use a GP emulator to approximate the likelihood function and, thus, carry out "approximate" Bayesian inference. This is done for the spatial ILM of Section 2.1 in an MCMC framework. We here follow the general setup of Jandarov et al. (2014) who used a GP emulator in the context of a spatial gravity model. The development of the emulator is carried out in two stages: (i) GP parameter estimation; and (ii) construction of the predictive distribution.

(i) GP parameter estimation. A GP is a stochastic process $\mathbf{D}(\mathbf{\Theta})$, over the space $\mathbf{\Theta} \in \mathbf{\Lambda}$, such that for any finite set $\mathbf{\Theta} = (\mathbf{\Theta}^{(1)}, \mathbf{\Theta}^{(2)}, \ldots, \mathbf{\Theta}^{(p)})^\top$, $\mathbf{D} = [D(\mathbf{\Theta}^{(1)}), D(\mathbf{\Theta}^{(2)}), \ldots, D(\mathbf{\Theta}^{(p)})]^\top$ has a multivariate normal distribution. Here $\mathbf{\Theta}^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_n^{(i)})$ is an input vector at the ith point in the input (ILM parameter) space, p is the number of points in the input space, and n is the number of input variables (ILM parameters). A GP is completely characterized by its mean and covariance functions, $\mu_{\mathbf{D}}(\mathbf{\Theta}) = \mathrm{E}[\mathbf{D}(\mathbf{\Theta})]$ and $K_{\mathbf{D}}(\mathbf{\Theta}^{(i)}, \mathbf{\Theta}^{(j)}) = \mathrm{Cov}[D(\mathbf{\Theta}^{(i)}), D(\mathbf{\Theta}^{(j)})]$, respectively.

Now let us suppose we have some observed data $(\mathbf{\Theta}^{(i)}, D(\mathbf{\Theta}^{(i)}))$, $i = 1, 2, \ldots, p$, and assume that the output values $D(\mathbf{\Theta}^{(i)})$ have been produced as a realization of a GP with mean and covariance functions $\mu_{\mathbf{D}}$ and $K_{\mathbf{D}}$. The mean function $\mu_{\mathbf{D}}$ is a linear function $\mu_D(\mathbf{\Theta}^{(i)}) = \beta_{0G} + \beta_{1G}\theta_1^{(i)} + \beta_{2G}\theta_2^{(i)} + \cdots + \beta_{nG}\theta_n^{(i)}$, where $\boldsymbol{\beta}_G = (\beta_{0G}, \beta_{1G}, \ldots, \beta_{nG})$ is the vector of unknown regression coefficients to be estimated. Many possible covariance structures exist. We here use the power law exponential family and define the ijth element of the covariance matrix by

$$K_{\mathbf{D}}(\mathbf{\Theta}^{(i)}, \mathbf{\Theta}^{(j)}) = \begin{cases} \tau_{GP}^2 \exp\left(-\sum_{k=1}^{n} \eta_k \left(\theta_k^{(i)} - \theta_k^{(j)} \right)^{\rho_k} \right), & \text{if } i \neq j, \\ \tau_{GP}^2 + \tau_\epsilon^2, & \text{otherwise,} \end{cases}$$

where $\tau_{GP}^2 = \mathrm{Var}[\mathbf{D}(\mathbf{\Theta})]$ is the unconditional variance of the GP; $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_n)$ is a vector of hyper parameters that control how far correlation extends in each input direction, playing the role of smoothing parameters; ρ_k are smoothness parameters that can be estimated but (as is typical) are here fixed at $\rho_k = 2$; and τ_ϵ^2 is a nugget parameter, representing variance due to response stochasticity (e.g., random noise).

In the context of our GP emulator, let $\Delta(\boldsymbol{\Theta}^{(i)}) = (\delta_1^{(i)}, \delta_2^{(i)}, \ldots, \delta_T^{(i)})$ be a set of T summary statistics of simulated data from our model (simulator) using parameters $\boldsymbol{\Theta}^{(i)}$ at the ith point in the parameter space. Further let $\mathbf{Z} = (z_1, z_2, \ldots, z_T)$ be the summary statistics of the observed data at unknown parameters $\boldsymbol{\Theta}^* = (\theta_1^*, \theta_2^*, \ldots, \theta_n^*)$, where z_t are T summary statistics of the observed data. Let us define the log L_2-norm distance between the summary statistics of simulated data generated under the ith parameter settings and the summary statistics of observed data to be $D(\boldsymbol{\Theta}^{(i)}) = \log\left[||\Delta(\boldsymbol{\Theta}^{(i)}) - \mathbf{Z}||^2\right]$ (here base 10 was used). Now under the GP emulator model we assume the distance vector $\mathbf{D} = [D(\boldsymbol{\Theta}^{(1)}), D(\boldsymbol{\Theta}^{(2)}), \ldots, D(\boldsymbol{\Theta}^{(p)})]^\top$ follows a multivariate normal distribution

$$\mathbf{D}|\boldsymbol{\Theta}, \boldsymbol{\beta}_G, \boldsymbol{\psi}_G \sim N(\mathbf{X}\boldsymbol{\beta}_G, \Sigma(\boldsymbol{\psi}_G)), \tag{3}$$

where $\mathbf{X}$ is a design matrix of dimension $p \times (n+1)$ whose ith row is $(1, \boldsymbol{\Theta}^{(i)})$; $\boldsymbol{\beta}_G = (\beta_{0G}, \beta_{1G}, \ldots, \beta_{nG})$ is a vector of regression coefficients; and $\boldsymbol{\psi}_G = (\tau_{GP}^2, \tau_\epsilon^2, \boldsymbol{\eta})$ is a vector of parameters of covariance matrix $\Sigma(\boldsymbol{\psi}_G) = [K_{\mathbf{D}}(\boldsymbol{\Theta}^{(i)}, \boldsymbol{\Theta}^{(j)})]$.

Now if the covariance matrix $\Sigma(\boldsymbol{\psi}_G)$ is fully specified the maximum likelihood estimates (MLEs) of $\boldsymbol{\beta}_G$ and τ_{GP}^2 exist and are given by $\hat{\boldsymbol{\beta}}_G = (\mathbf{X}^\top \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Sigma^{-1} \mathbf{D}$ and $\hat{\tau}_{GP}^2 = \frac{1}{p}\left(\mathbf{D} - \mathbf{X}\hat{\boldsymbol{\beta}}_G\right)^\top \Sigma^{-1} \left(\mathbf{D} - \mathbf{X}\hat{\boldsymbol{\beta}}_G\right)$, respectively.

(ii) Construction of the predictive distribution. Let $\hat{\boldsymbol{\beta}}_G = (\hat{\beta}_{0G}, \hat{\beta}_{1G}, \ldots, \hat{\beta}_{nG})$ and $\hat{\boldsymbol{\psi}}_G = (\hat{\tau}_{GP}^2, \hat{\tau}_\epsilon^2, \hat{\boldsymbol{\eta}})$ be the MLEs of $\boldsymbol{\beta}_G$ and $\boldsymbol{\psi}_G$, respectively. We here are interested in predicting output $D^* = \log\left[||\Delta(\boldsymbol{\Theta}^*) - \mathbf{Z}||^2\right]$, where $\Delta(\boldsymbol{\Theta}^*)$ is a new simulated data set at unknown parameter $\boldsymbol{\Theta}^*$. Let $\boldsymbol{r} = (r_1, r_2, \ldots, r_i, \ldots, r_p)$, where $r_i = \text{cor}(D(\boldsymbol{\Theta}^*), D(\boldsymbol{\Theta}^{(i)}))$. Then by multivariate normal theory (e.g., Anderson, 1984), D^* follows a normal distribution for given $\mathbf{D}$, $\boldsymbol{\Theta}$, and $\boldsymbol{\Theta}^*$. Specifically the predictive distribution for D^* is given by

$$D^*|\mathbf{D}, \boldsymbol{\Theta}, \boldsymbol{\Theta}^* \sim N\left(\mu^*, \Sigma^*\right), \tag{4}$$

where $\mu^* = \hat{\beta}_{0G} + \theta_1^* \hat{\beta}_{1G} + \theta_2^* \hat{\beta}_{2G} + \cdots + \theta_n^* \hat{\beta}_{nG} + \hat{\tau}_{GP}^2 \boldsymbol{r}^\top (\Sigma(\hat{\psi}_G))^{-1} (\mathbf{D} - \mathbf{X}\hat{\boldsymbol{\beta}}_G)$ and $\Sigma^* = \hat{\tau}_{GP}^2 + \hat{\tau}_\epsilon^2 - \hat{\tau}_{GP}^4 \boldsymbol{r}^\top (\Sigma(\hat{\psi}_G))^{-1} \boldsymbol{r}$. We denote this predictive distribution by $f_{GP}(D^*|\boldsymbol{\Theta}^*)$. Further detail of the derivation of this predictive distribution can be found in Santner, William, & Notz (2003).

2.3. Model Fitting with the GP Emulator

We initially consider summary statistics $\boldsymbol{\Delta}$ (and $\mathbf{Z}$) based on simulated (and observed) epidemic curve data across a whole population (in Section 2.4 we generalize this to allow for a comparison between a series of spatially stratified epidemic curves). We specifically let $\delta_t^{(i)}$ be the number of I individuals at time point t from simulated data set i (with the ith parameter setting), and z_t be the number of I individuals at time point t in the observed data. Thus the number of summary statistics is the number of days of the observed epidemic (i.e., $T = t_{max}$) and $\boldsymbol{\Delta}$ and $\mathbf{Z}$ represent the simulated and observed epidemic curves, respectively.

We generally would expect the distance $\mathbf{D}(\boldsymbol{\Theta})$ between $\boldsymbol{\Delta}$ and $\mathbf{Z}$ to be small—indicating good fit—in regions of the parameter space where the likelihood is high. Thus to carry out quick approximate inference the basic notion is to substitute the likelihood $f(\mathscr{D}|\theta)$ with the probability of observing a distance $D^* = \lambda$ under our GP likelihood approximation in line with the observed data (i.e., the GP predictive distribution $f_{GP}(D^* = \lambda|\boldsymbol{\Theta}^*)$). We naively might consider $\lambda = 0$, although this turns out not to be optimal for reasons explained below.

There will be three sources of error in this approximation. The first source of error results from any lack of sufficiency in the summary statistics being used to summarize the data. As finding

sufficient statistics is not typically possible for complex models, our aim is to find statistics with as high a sufficiency level as possible without compromising computational tractability. If sufficient statistics can be found for a particular problem, this source of error will, of course, be negated.

The second source of error is the Monte Carlo error introduced via the epidemic simulation. The ILM defines a stochastic process and, thus, an infinite number of simulations would be required to negate this error. The final source of error results from the approximation of the GP to the likelihood surface. The function $f_{GP}(D|\boldsymbol{\Theta})$, although much faster to compute than the likelihood, is likely to be far less flexible than the likelihood function given its underlying Gaussian structure.

(i) GP discrepancy term. These final two sources of error are known collectively as the "data-model discrepancy." Kennedy & O'Hagan (2001) introduce such a discrepancy term to represent model inadequacy. As pointed out by Bayarri et al. (2007), ignoring the data-model discrepancy often leads to a poorer inference, and so it is common to treat it as an unknown parameter in the analysis.

Following Jandarov et al. (2014) we denote the discrepancy $D^* = \lambda, \lambda > 0$ and then construct a predictive distribution $f_{GP}(\lambda|\boldsymbol{\Theta}^*)$ incorporating λ as an additional unknown parameter. The prior for the discrepancy term λ needs to be reasonably carefully chosen (Bhat, Haran, & Goes, 2010) and is generally chosen to be somewhat informative in order to avoid problems with identifiability (and, thus, MCMC performance). In fact plug-in point estimates are often used (Bhat, Haran, & Goes, 2010). As λ is a term representing the discrepancy between the emulator and ILM likelihood, the average log L_2-norm distance between two epidemic curves with the same parameter values, estimable via simulation, should provide a rough lower bound for λ that can be used to inform prior choice for this parameter.

(ii) Emulator-based MCMC. Once the design matrix has been determined detailing for which values of the parameters simulations will be carried out, $D^{(i)} = \log\left[||\Delta(\boldsymbol{\Theta}^{(i)}) - \mathbf{Z}||^2\right]$ is calculated for each i. The GP is fitted, taking $\mathbf{D} = (D^{(1)}, D^{(2)}, \ldots, D^{(p)})$ as response variable and the ILM parameters (α and β) as the predictors, and maximum likelihood estimates of GP parameters obtained. We then construct the predictive distribution of the GP parameters.

The ILM is then fitted to the observed epidemic data using the predictive distribution $f_{GP}(\lambda|\alpha, \beta)$ under the GP emulator as an approximation of the likelihood function of the ILM in the Bayesian setting. The joint posterior distribution for α and β (and λ) up to a constant of proportionality is approximated by

$$\pi_{GP}(\alpha, \beta, \lambda|\mathbf{D}) \propto f_{GP}(\lambda|\alpha, \beta)\mathrm{p}(\alpha)\mathrm{p}(\beta)\mathrm{p}(\lambda), \tag{5}$$

and this is sampled from using an MCMC algorithm (see Section 3.1).

2.4. Spatially Stratified Epidemic Curves

We refer to the use of epidemic curve data across the entire population to form $\mathbf{D}$ as using a global epidemic curve. One of our objectives here is also to compare the use of global-based emulators with those based on spatially stratified epidemic curve data, as successfully used for epidemic curve classification (Pokharel & Deardon, 2014). Thus consider a population spatially split into s strata. We denote the simulated and observed epidemic curves recorded in the kth stratum to be $\Delta_k(\boldsymbol{\Theta}^{(i)}) = (\delta_{k1}^{(i)}, \delta_{k2}^{(i)}, \ldots, \delta_{kt_{max}}^{(i)})$ and $\mathscr{Z}_k = (z_{k1}, z_{k2}, \ldots, z_{kt_{max}})$, respectively, where $\delta_{kt}^{(i)}$ and z_{kt} are the number of I individuals in stratum k at time t generated for the ith parameter setting for simulated epidemics, and observed epidemics, respectively. Now the full epidemic curve data sets are given by $\bar{\Delta}(\boldsymbol{\Theta}^{(i)}) = (\Delta_1(\boldsymbol{\Theta}^{(i)}), \Delta_2(\boldsymbol{\Theta}^{(i)}), \ldots, \Delta_s(\boldsymbol{\Theta}^{(i)}))$ and $\bar{Z}(\boldsymbol{\Theta}^*) = (\mathscr{Z}_1, \mathscr{Z}_2, \ldots, \mathscr{Z}_s)$ for simulated and observed epidemics, respectively. Further our distance metric for each i becomes

$D^{(i)}(\boldsymbol{\Theta}^{(i)}) = \log\left[||\bar{\Delta}(\boldsymbol{\Theta}^{(i)}) - \mathbf{Z}||^2\right]$ for each i. Following Pokharel & Deardon (2014), we use two variants of spatial stratification: circular and rectangular.

(i) Circular stratification. In this method we construct concentric circles of increasing radius centred at the initial infection. We thus generate the strata as the disjoint rings of the circle with equal width. The smallest circle would therefore be the central stratum, and the other strata would consist of rings, some of which, in our case, being truncated by the edge of the rectangular population observed.

(ii) Rectangular stratification. The global population is here divided into several rectangular blocks. Two forms of this method are considered here. In the first, regular rectangular stratification, rectangular strata of equal size are formed by imposing a regular grid of rectangles over the global population. In the second, irregular rectangular stratification, the same type of grid is imposed upon the global population to form the strata, but the grid is shifted to ensure that the first infection is approximately at the centre of one of the inner strata. It is here important to note that shifting the grid lines to make sure the first infection lies approximately at the centre of one of the strata leads to strata of different sizes over the global population of interest.

2.5. Posterior Predictive Statistics

We use a posterior predictive approach for model assessment, comparing summary statistics of observed epidemic data to the posterior predictive distribution of said statistics under the fitted model (Gelman, Meng, & Stern, 1996).

The primary summary statistic considered is the epidemic curve itself which is, of course, a global summary statistic of the epidemic containing no information about spatial dynamics. We therefore also consider the posterior probability distribution of the probability of one-to-one infection curve over varying distance as used in Kwong & Deardon (2012). This curve is obtained by imagining that $I(t)$ contains only one individual. Thus in terms of the general ILM of Equation (1), it is given by considering the effect of varing d_{ij} upon

$$P_{it} = 1 - \exp\left[-\alpha d_{ij}^{-\beta}\right];\ \alpha, \beta > 0.$$

3. DATA ANALYSIS

3.1. Simulated Data Analysis

A population of 625 individuals was created with the X and Y coordinates of individuals generated from a bivariate normal distribution, such that

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim MVN\left(\begin{pmatrix} 12.5 \\ 12.5 \end{pmatrix}, \quad \begin{pmatrix} 8.0 & 0 \\ 0 & 8.0 \end{pmatrix}\right).$$

Figure 1 shows the population generated using this bivariate normal distribution.

Ten epidemics were simulated across the population using the power-law spatial ILM described in Section 2 with model parameters α, β, and γ_I. We considered a fixed known value of the I period of $\gamma_I = 2$. The values of parameters α and β used to simulate an observed epidemic were 0.2 and 2.5, respectively. The reason for choosing these parameter values was to produce informative epidemic data (i.e., epidemics that neither quickly die out nor very quickly infect the whole population). One individual approximately at the centre was set as the initial infection for each simulated epidemic at time point $t = 1$. The epidemics were then propagated via the spatial ILM. The epidemics ran from $t = 1$ to t_{max}, where t_{max} is the time when either the infection dies out or the whole population becomes infected.

The design matrix for constructing the emulator was defined by a uniform grid $\alpha \in [0.1, 1.0]$ and $\beta \in [2.1, 3.0]$. To evaluate the robustness of the emulation-based analysis, four versions of the grid were used, containing 10, 15, 20, and 25 points in the parameter interval. Thus 100, 225, 400, or 625 epidemics were generated using 10^2, 15^2, 20^2, or 25^2 combination of parameters in the respective grids. The relevant log L_2-norm distance $\mathbf{D}$ between summary statistics of observed and simulated epidemic data was calculated (i.e., based upon global or spatially stratified epidemic curves). Both full Bayesian and emulator-based models were fitted to each of the 10 epidemic realizations, emulation-based models being fitted to both global and each type of spatially stratified epidemic curve data.

Various degrees of circular stratification resolution were considered, with strata formed from 2, 3, 4, 5, 6, or 7 concentric circles with equal ring width, giving central circular strata with radii of 8.96, 5.97, 4.48, 3.58, 2.98, or 2.56 units, respectively, and each other stratum defined by increasingly larger rings of the same respective widths.

Rectangular stratifications under 2×2, 3×3, 4×4, and 5×5 degrees of resolution (i.e., 4, 9, 16, or 25 strata) were considered. Of course in an $n \times n$ rectangular stratification scheme where $n = 2z + 1$, $z \in \mathbb{Z}^+$, if the initial infection lies in the centre of a square population (e.g., the 5×5 resolution) then the regular and irregular rectangular stratification would be equivalent. As the degree of resolution increases (i.e., as the number of strata increase), there is also less and less difference between regular and irregular stratification. However for resolutions such as 2×2 and 4×4, the two forms of rectangular stratification can be quite different. See Figure 1, for examples, of both circular and rectangular stratification.

Independent, non-informative exponential priors for model parameters α and β, with mean values of $\xi_1 = \xi_2 = 10^5$, were used. A more informative exponential prior with mean $\xi_3 = 10$ was used for the data-model discrepancy parameter λ. The value of ξ_3 was chosen after observing the distribution of Δ for a few simulated epidemics (see Section 2.3). Further a small simulation study using values of ξ_3 ranging from 1 to 10^5 was carried out and $\xi_3 = 10$ was found to work well (e.g., in terms of MCMC mixing). In fact both parameter estimates and predictive performance were observed not to be particularly sensitive to choices of ξ_3.

Random walk Metropolis Hastings MCMC chains were run for 200,000 iterations with a burn-in period of 10,000. This was done for both the full Bayesian and the emulation-based analyses. Random-walk updates were tuned and convergence was confirmed through visual inspection. For example, for global epidemic curve-based GP emulation, updates were based on $U[-0.01, 0.01]$, $U[-0.1, 0.1]$, and $U[-10, 10]$ noise being added to current values to give candidate values the for α, β, and λ parameters, respectively. Posterior mean values and 95% percentile interval (PI) for each model parameter and the discrepancy term were calculated. The CPU time to run the MCMC chain was recorded for all analyses, using a 12 core (2×2.93 GHz 6-core Intel Xeon), 12 GB RAM Mac Pro. The MCMC algorithm was coded in GNU Fortran. The GP emulator was constructed using the mlegp R package of Dancik & Dorman (2008).

3.2. TSWV Data Analysis

Our second data set is from an experiment on TSWV, as described in Hughes et al. (1997). They describe an experiment in which disease assessments were made at regular time intervals on a total of 520 pepper plants grown inside a greenhouse. Plants were placed in 26 rows spaced one metre apart, with 20 plants spaced half a metre apart in each row; hence, the plants were located at points (x_i, y_i), $x_i = 1, 2, \ldots, 26$ and $y_i = 0.5, 1.0, \ldots, 9.5, 10.0$. The experiment began on May 26, 1993 when the first disease assessment occurred. Subsequent biweekly assessments took place from June 9 to August 16, 1993. Therefore for our discrete-time model framework, May 26 was set as $t = 1$, June 9 was set as $t = 2$, up to the last observation on August 16, 1993 as $t = 7$, where each time increment was 14 days. A total of 327 infected plants were recorded during the 82 days

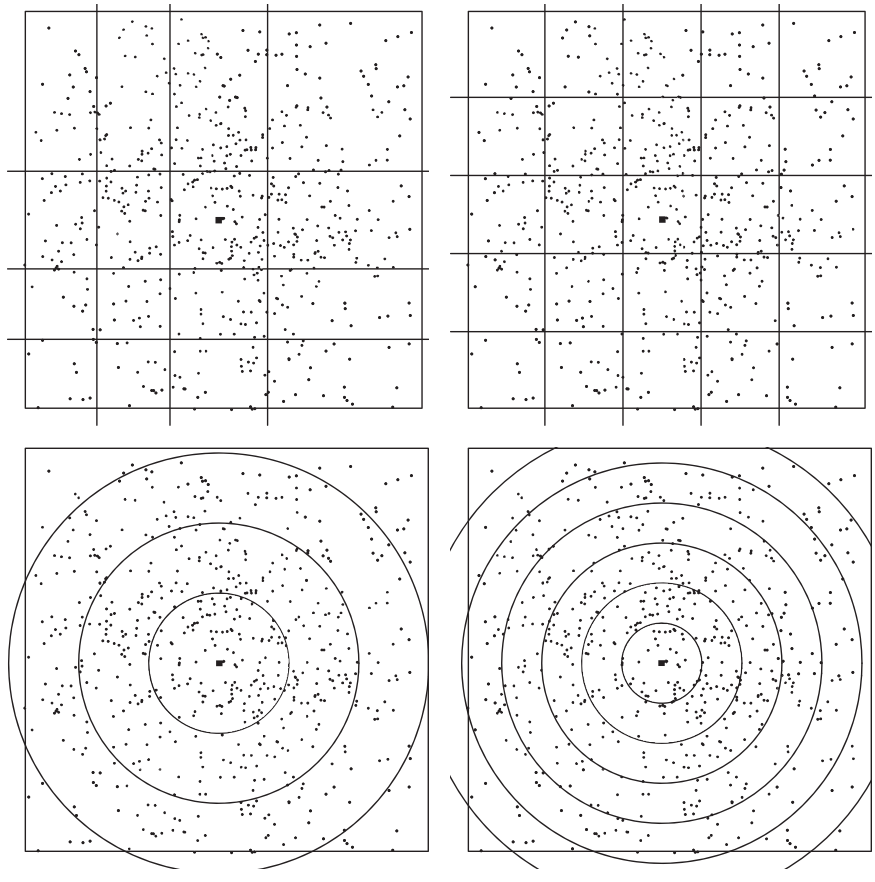

FIGURE 1: Rectangular and circular spatial stratifications of the global population. The small black square approximately at the centre of the population is an I individual at time point $t = 1$. The left top panel represents 4×4 irregular stratification and the top right panel represents 5×5 regular stratification. The bottom left and right panels represent circular stratification of four and seven rings of equal width, respectively.

of the experimental period. In our analysis we set the I period to be $\gamma_I = 3$ (i.e., 42 days), which is in line with Brown et al. (2005) who indicate that the estimated I period of a TSWV epidemic is 6 weeks.

The design matrices were chosen based on results from Pokharel & Deardon (2014), in which random forest-based classifiers were used to identify parameters in a power-law spatial ILM for this data set. The intervals therefore considered for the two model parameters were $\alpha \in [0.005, 0.5]$ and $\beta \in [1.0, 2.0]$. Once again a 20^2 grid was found adequate for constructing a GP emulator. Following the same procedure used for the simulated data sets, both the full Bayesian and the emulation-based models were fitted to both global and spatially stratified epidemic curve data.

As in the case of simulated data we considered strata formed from 2, 3, 4, 5, 6, or 7 concentric circles with equal ring width, giving central circular strata with radii of 11.24, 7.49, 5.62, 4.49, 3.75, or 3.21 units, respectively. For the case of rectangular stratification, we considered both forms of stratification (regular and irregular) under 2×2, 3×3, 4×4, and 5×5 degrees of resolution (i.e., 4, 9, 16, or 25 strata).

Non-informative exponential priors with means $\xi_1 = \xi_2 = 10^5$ were placed upon model parameters α and β. An informative exponential prior Exp(ξ_3) with mean $\xi_3 = 100$ was used for the data-model discrepancy term λ. Once again the value of ξ_3 was chosen after observing the

distribution of Δ for a few simulated epidemics. A small simulation study using values of ξ_3 ranging from 1 to 10^5 was carried out and $\xi_3 = 100$ was found to work well. The same computer and coding setup was used as for the simulated data analysis, and the CPU time taken to run various MCMC analyses was recorded.

4. RESULTS

4.1. Results for Simulated Data

Two-dimensional posterior plots for (α, β) obtained for the emulation-based ILM analysis over different parameter grids for a typical epidemic realization, as well as equivalent plots for the full Bayesian analysis, are given in Figure 2. These plots clearly show that the posterior surface plot for the emulation-based results approximately coincides with that of the full Bayesian analysis. However parameter grid size has some effect on the quality of the results. For lower resolutions (e.g., grid sizes of 10^2 and 15^2), more biased inference was observed relative to that under the higher resolutions (e.g., grid sizes of 20^2 and 25^2), although the 95% posterior ellipse under all analyses except the 10^2 grid resolution did capture the true parameter values of $\alpha = 0.2$ and $\beta = 2.5$. Overall the 20^2 parameter grid appeared to provide the best approximation to the true posterior under a global epidemic curve-based emulator. In fact for the 20^2 grid, the 95% PIs captured the true value of both parameters in eight of the ten epidemic realizations, the exceptions

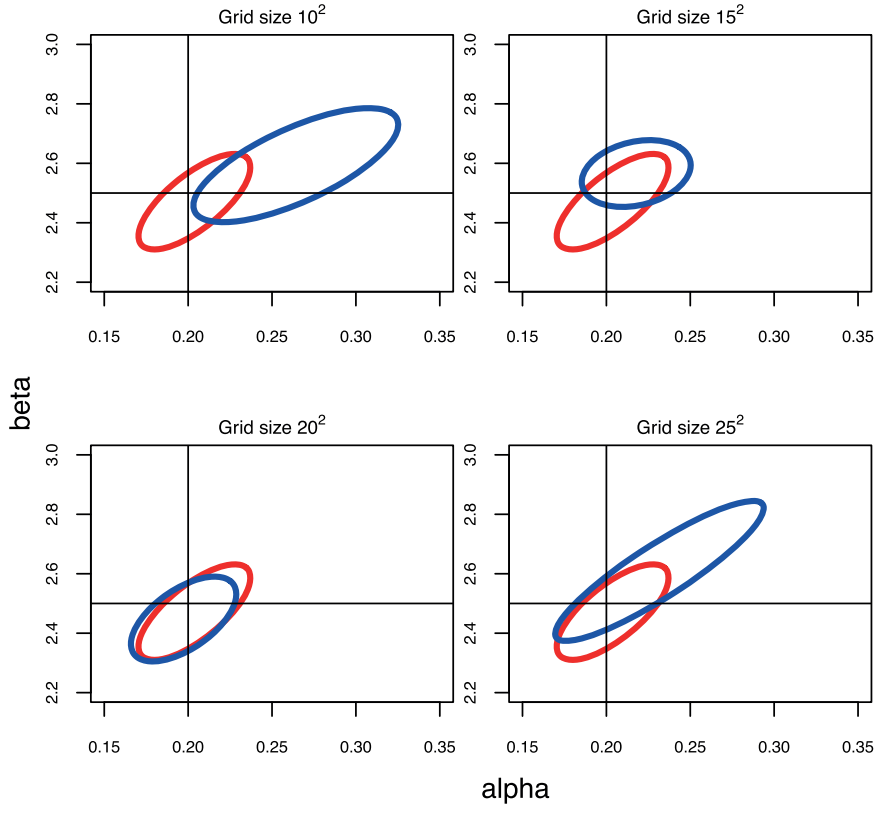

FIGURE 2: Ninety-five percent approximate posterior ellipses for α and β for one typical simulated epidemic with emulators based upon global epidemic curves for different design matrix resolutions. The red ellipses are the true 95% posterior ellipses and the blue ellipses are the emulation-based 95% approximate posterior ellipses. The black lines represent the true values for the parameters.

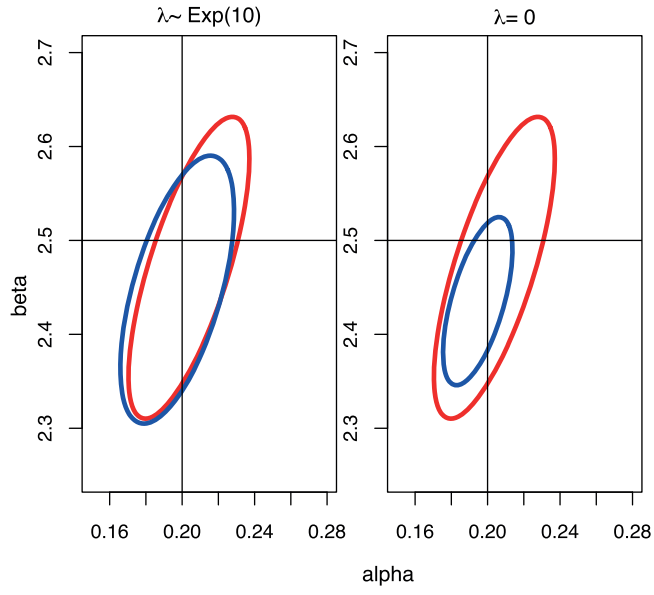

FIGURE 3: Ninety-five percent approximate posterior ellipses for α and β for one typical simulated epidemic with emulators based upon global epidemic curves with and without the discrepancy term for 20^2 grid resolution. The red ellipses are the true 95% posterior ellipses and the blue ellipses are the emulation-based 95% approximate posterior ellipses. The black lines represent the true values for the parameters.

being one wherein α, and one wherein β, was not captured. Posterior plots for the other epidemic realizations were similar (results not shown).

We also fitted the emulation-based model with no discrepancy term (i.e., $\lambda = 0$). Two-dimensional posterior surface plots for a typical emulation-based, and full Bayesian, analysis based upon the global epidemic curves using a 20^2 parameter grid size are given in Figure 3. We can see that the approximate posterior mean was close to the full Bayesian posterior mean, with both capturing the true parameter values, but with a much tighter approximate posterior credible ellipse when $\lambda = 0$. These results, which were again similar for other epidemic realizations, suggest that adding a discrepancy term may help prevent a distinct underestimation of posterior uncertainty. We note, however, that they are in contrast with those of Bayarri et al. (2007) and Jandarov et al. (2014) who suggest a narrowing of model parameter PIs when the discrepancy term is used.

Figure 4 shows a typical emulation-based posterior approximation produced using spatially stratified epidemic data (circular stratification). Results for rectangular stratifications were similar and are not included, although overall circular stratification was seen to offer slightly better inference than rectangular stratification. Overall, however, the results obtained for the emulator built using the spatially stratified epidemic curves were disappointing. Although performance was reasonably good for some stratification settings (e.g., circular stratification with four rings and a 20^2 parameter grid), it is hard to argue that there was noticeable improvement in the posterior approximation even for the best stratification schemes. Further poorly selected stratification schemes appeared to introduce a large degree of posterior bias and/or imprecision. We finally note that the use of irregular versus regular rectangular stratification was seen to have little effect on performance (results also not shown).

Figure 5 displays the mean and 95% PIs of the posterior predictive distribution of the epidemic curve, for both full Bayesian analysis and emulation-based analysis for emulators built under the various grid resolutions of the parameter space for one typical epidemic realization of 10

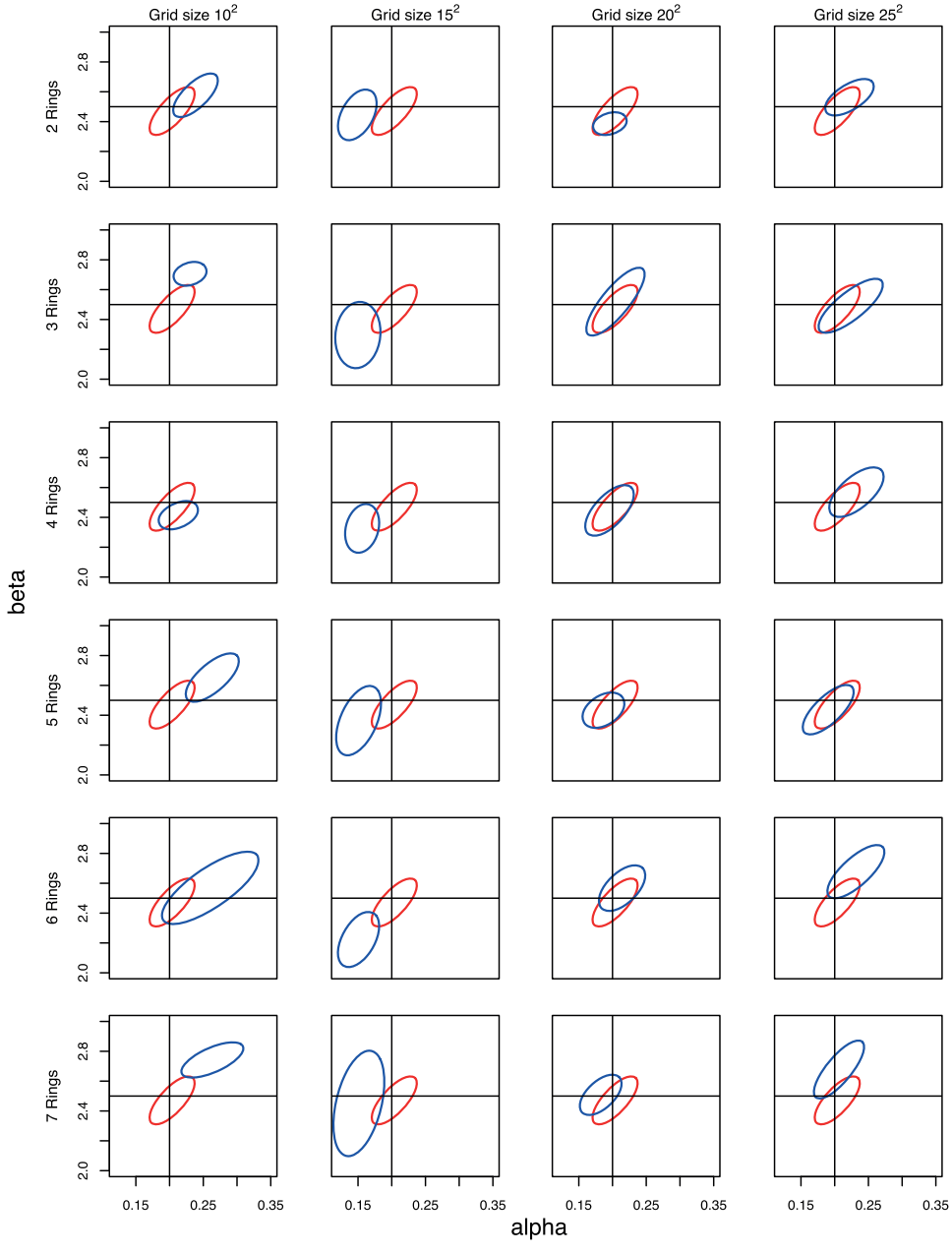

FIGURE 4: Ninety-five percent approximate posterior ellipses for α and β for one typical simulated epidemic with emulators based upon circularly stratified epidemic curves. Red ellipse represents the true posterior 95% PI and blue ellipse represents the emulation-based 95% PI. The black lines represent the true values for the parameters.

realizations produced. The emulation-based inference produced very similar results to the full Bayesian analysis regardless of the grid resolution used to build the emulator. Figure 6 shows the posterior predictive one-to-one probability curve for the same epidemic realization. Once again the emulation-based model produced the results very similar to that of the full Bayesian model,

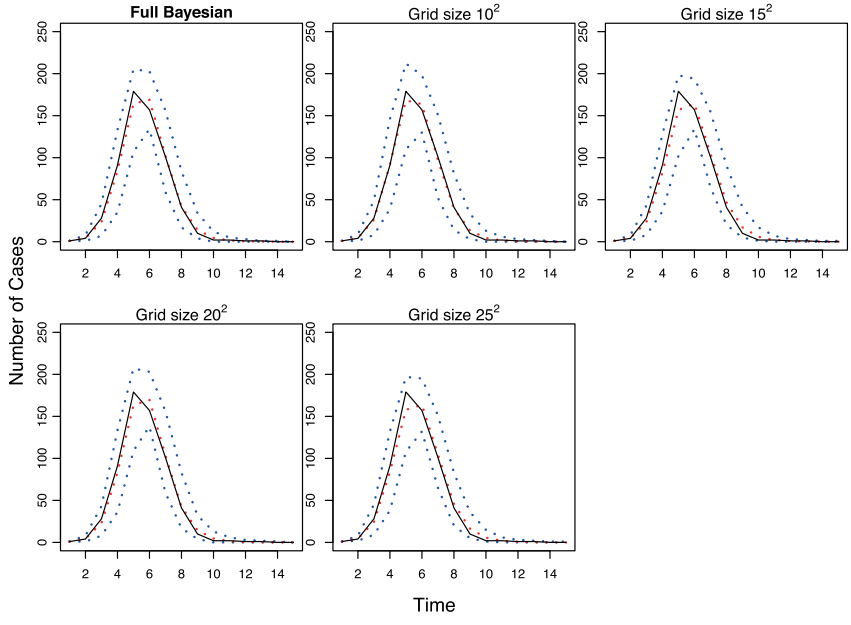

FIGURE 5: Posterior predictive mean epidemic curve (….), 95% PI (….) and true epidemic curve (___) for one typical simulated epidemic.

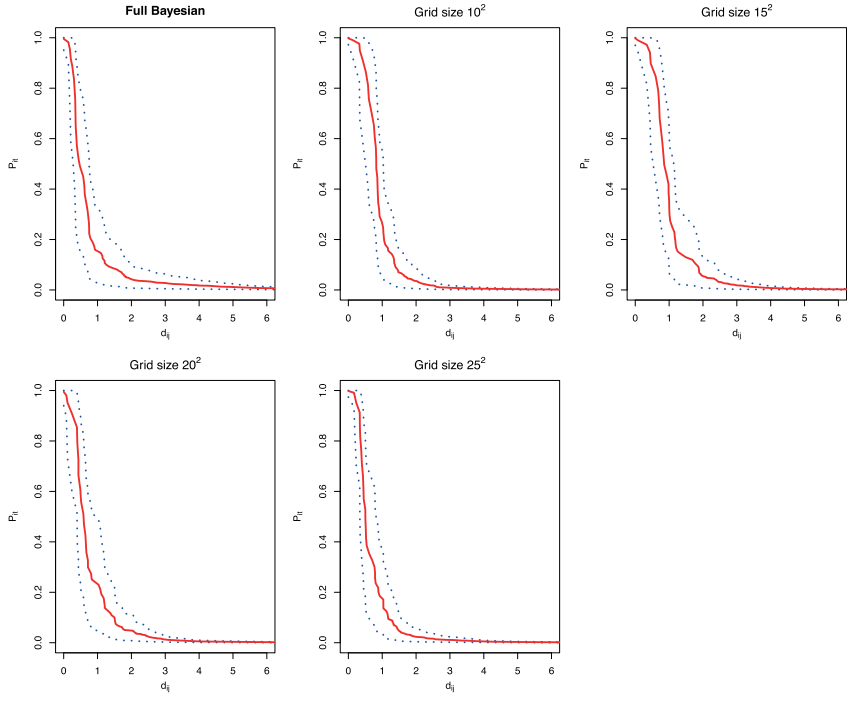

FIGURE 6: Posterior predictive mean one-to-one infection probability (___) and 95% PI (….) for one typical simulated epidemic.

TABLE 1: Mean CPU time to run 200,000 MCMC iterations for both the full Baysian and emulation-based models with different grid sizes in the parameter space.

Model	Grid size	Time (seconds)
Bayesian model	–	3,533.69
Emulation-based model	10	13.99
	15	57.47
	20	168.15
	25	392.61

and these results were reasonably invariant to grid resolutions. Similar posterior predictive results were observed for the remaining epidemic realizations (results not shown).

The CPU time taken to run 200,000 MCMC iterations for both the full Bayesian and emulation-based models, averaged over the 10 epidemic realizations, are given in Table 1. The full Bayesian analysis was much more computationally expensive than the emulation-based methods. For example, the emulation-based method was roughly 20 times faster using the 20^2 parameter grid to build the GP emulator than the full Bayesian method. Note that increasing the resolution of the parameter grid leads to slower performance of the emulation-based methods. This is due to the increased cost in calculating the predictive distribution. Overall a 20^2 parameter grid seemed to be most appropriate considering both computational and inferential performance.

4.2. Results for TSWV Data

The posterior estimates of α, β, and the data-model discrepancy term λ for both the full Bayesian and emulation-based methods are shown in Table 2 and Figure 7. Posterior mean estimates of model parameters for the emulation-based models were generally close to the true posterior estimates, implying the emulation-based model captures the biological characteristics of the spatio-

TABLE 2: Estimates of model parameters and the data-model discrepancy term using the full Bayesian model and a GP emulator for TSWV epidemic data.

Model	Grid resolution	Parameter estimate	α	β	λ
Bayesian model	–	Mean	0.0194	1.3597	–
		95% PI	(0.0092, 0.0296)	(0.8567, 1.7553)	–
Emulator model	10^2	Mean	0.0494	1.7156	5,078.292
		95% PI	(0.0395, 0.0620)	(1.6837, 1.7497)	(4,120.627, 6,058.227)
	15^2	Mean	0.0328	1.4256	3,542.914
		95% PI	(0.0233, 0.0427)	(1.3465, 1.5305)	(2,660.207, 4,566.028)
	20^2	Mean	0.0227	1.2964	4,258.059
		95% PI	(0.0166, 0.0287)	(1.2023, 1.4886)	(3,273.954, 5,334.627)
	25^2	Mean	0.0197	1.0504	2,886.655
		95% PI	(0.0119, 0.0263)	(0.8995, 1.1033)	(1,967.181, 3,954.896)

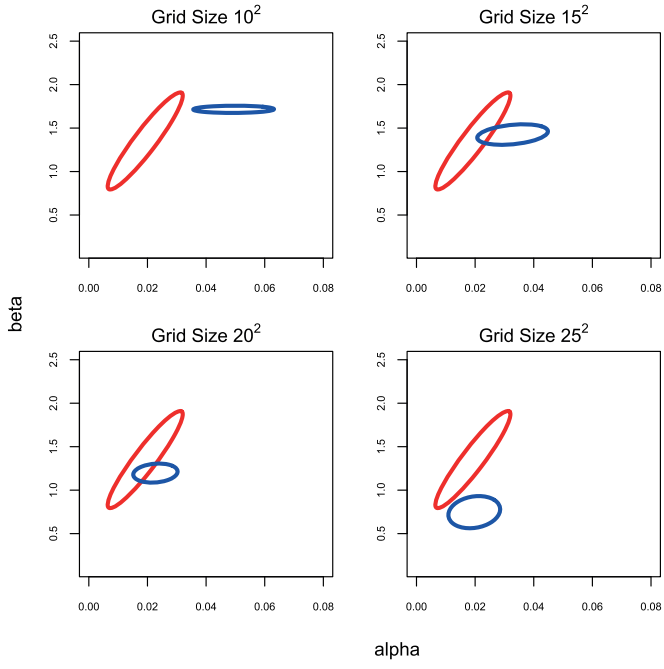

FIGURE 7: Ninety-five percent approximate posterior ellipses for α and β for TSWV epidemic with emulators based upon global epidemic curves for different design matrix resolutions. The red ellipses are the true 95% posterior ellipses and the blue ellipses are the emulation-based 95% approximate posterior ellipses.

temporal I disease system fairly well. The posterior mean estimates of α and β were also fairly close to true posterior estimates under different grid resolutions, but that too low (10^2) or too high (25^2) a grid resolution produced results that were not as good as those for grid resolutions of 15^2 and 20^2; this result is in line with the results of simulated data. Furthermore the authors found that increasing the grid resolution beyond this (e.g., to 30×30) failed to produce any discernible improvement in parameter estimation.

In all cases, especially under the 10^2 resolution design matrix, the posterior uncertainty in β was distinctly underestimated. However in terms of predictive performance the emulator-based models tended to do well and be fairly robust to grid size. For example, Figure 8 shows the probability of the one-to-one infection curve over distance for the TSWV data set under the full Bayesian and global curve-based emulator models. We can see a slightly more marked effect from the grid resolution chosen than was observed for the simulated data, especially for low distances. However the results from the 20^2 grid resolution appeared to be closely matched with the results from the full Bayesian analysis.

Figure 9 shows the approximate posterior results obtained under rectangular spatial stratification (results obtained under circular stratification were similar). Spatial stratification did not appear to improve the posterior mean estimates. Generally estimation of the infectivity parameter α worsened, and estimation of the spatial parameter β improved as the resolution of the stratification increased. Good MCMC chain mixing also got progressively harder to achieve (especially for the spatial parameter β) as the stratification resolution increased. However under the higher resolutions of spatial stratification the approximate posterior uncertainty was found to greatly increase, thus, reducing the problem of uncertainty underestimation observed with the global epidemic curve-based emulator.

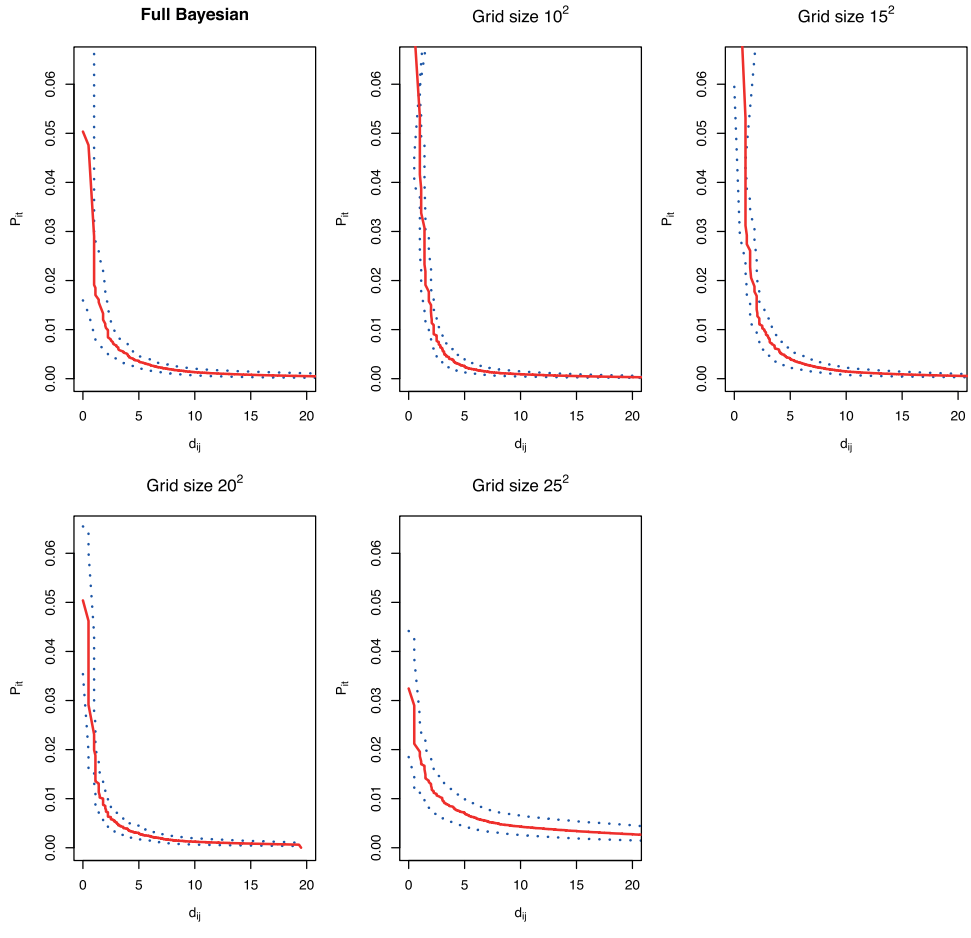

FIGURE 8: Posterior predictive mean one-to-one infection probability (——) and 95% PI (....) for TSWV analysis.

In terms of the computation time the global curve-based full Bayesian analysis was much more highly computationally expensive (6,834 s to run 200,000 iterations) compared to the emulation-based methods (166 s to run 200,000 iterations) using the 20^2 parameter grid to build the GP emulator. This is around 41 times higher. The results for the other grid resolutions were similar to the simulated data and not are included here.

We finally note that the discrepancy posterior was found to have high mass far from the central mass of the prior, seemingly implying a poor choice of prior for this parameter. In fact as was the case for the simulated data analyses, both parameter estimates and predictive performance were not observed to be sensitive to the choice of prior for the discrepancy term (as long as the discrepancy prior is not very tightly packed at zero).

5. DISCUSSION

We have used both full Bayesian and emulation-based MCMC methods to analyze simulated data and data from an experiment on TSWV. Our primary objective here was to see how well emulation-based methods perform relative to full Bayesian analysis at capturing the biological characteristics

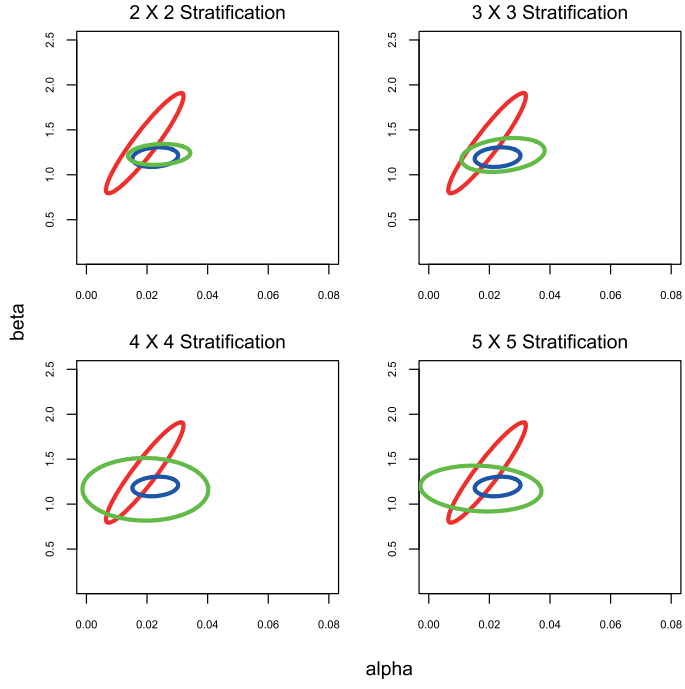

FIGURE 9: Ninety-five percent approximate posterior ellipses for α and β for TSWV epidemic with full Bayesian (____), and emulators based upon global epidemic curves (____), and rectangular stratified epidemic curves (____).

of small spatial infectious disease systems as a proof-of-concept. Due to the GP assumption of stationarity of the covariance function and common variance parameter τ^2_{GP}, nonstationarity in the simulator (here the ILM likelihood) can result in poor estimation (Bastos & O'Hagan, 2009). However our results imply that emulation-based methods can infer the biological characteristics of at least some spatial infectious disease systems fairly well, and they certainly do offer a much quicker mode of analysis than the full Bayesian MCMC analysis.

In addition to the simulated and real data scenarios described, many other simulated scenarios have been tested by the authors that have not been included for brevity. For example, a number of simulation studies equivalent to the one described in Section 3.1 but for different parameter values have been carried out. These studies tended to produce similar results to those described in Section 4.1 in cases wherein the true parameter values are captured within the span of the chosen design matrix; that is, posterior credible intervals that capture the true parameter values, and posterior predictive distributions that contain the observed epidemic curve or one-to-one infection probability over varying distance, are generally observed.

However emulator-based analyses wherein the true parameter values are not contained within the span of the chosen design matrix have been seen to produce somewhat different results. For example, consider a scenario using the same uniform grid over $\alpha \in [0.1, 1.0]$ and $\beta \in [2.1, 3.0]$ with a 20^2 resolution as considered in Section 3.1, but with parameters $\alpha = 0.8$ and $\beta = 1.5$; a case in which the spatial parameter, β does not lie within the span of the design matrix. Under this scenario, it was observed that the true value of α was captured within the 95% PI for each of 10 simulated epidemic data sets, but that β was captured within none. Furthermore predictive ability was substantially poorer. For example, the posterior predictive epidemic curve distributions consistently failed to capture the true epidemic curve for the 10 data sets under this scenario

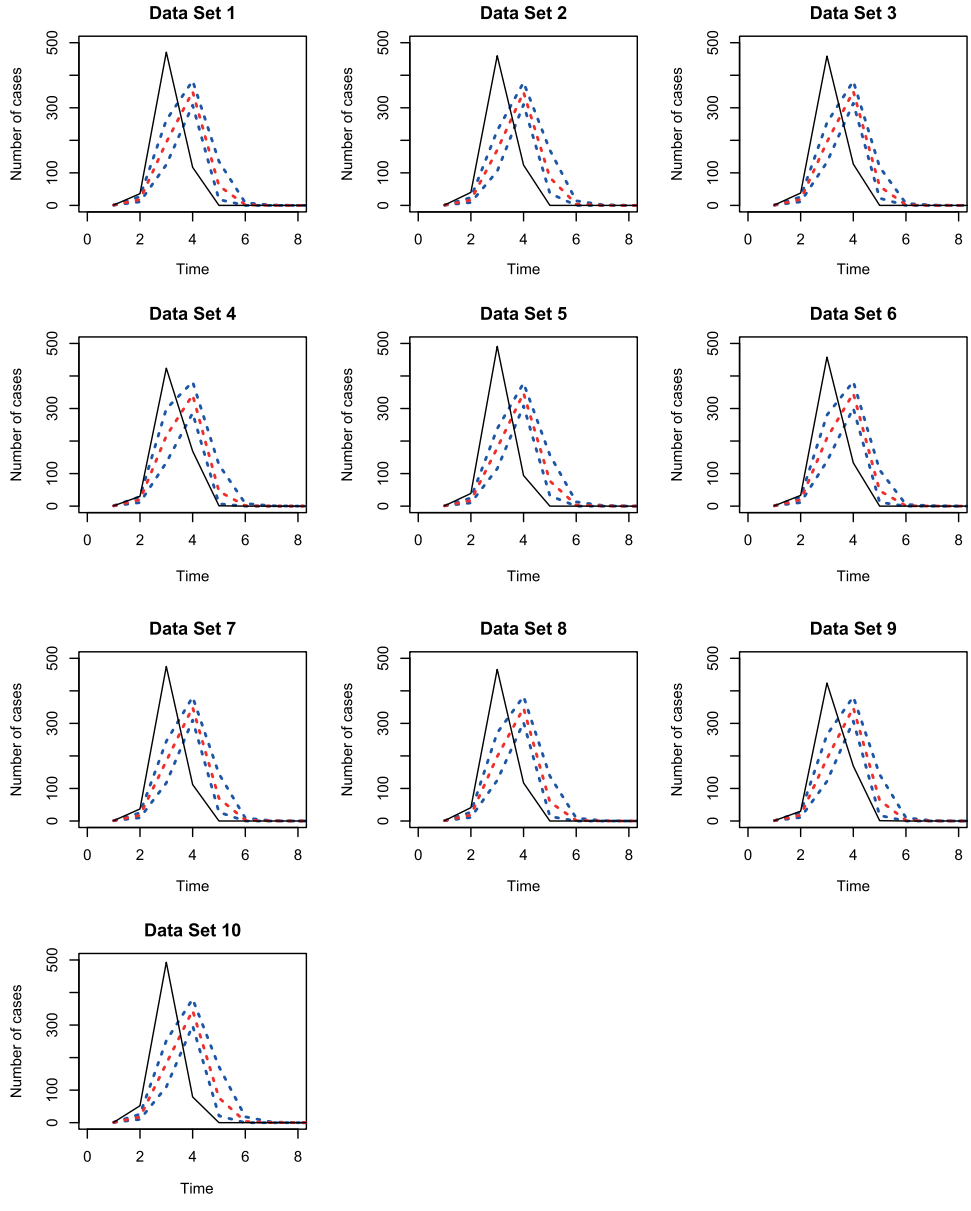

FIGURE 10: Posterior predictive mean epidemic curve (....), 95% PI (....), and true epidemic curve (____) for one epidemic realization in which parameters ($\alpha = 0.8$ and $\beta = 1.5$) are not captured within the span of the design matrix.

(Figure 10). Similar results were observed in other situations where the span of the design matrix failed to capture the true parameter values.

This implies that such approximate posterior predictive plots could be used diagnostically to help determine a suitable design matrix span, although the results of Section 4.1 suggest such plots do not aid much, for example, in the choice of grid resolution once the span of the design matrix has been suitably set. The extent of the robustness of such approaches to aid in design matrix formulation, which summary statistics lead to the best diagnostics for such tasks, and possibly

which summary statistics are best used for building the emulator, is something the authors intend to explore in future work.

In terms of both parameter estimation and computation time there was a noticeable effect on parameter estimation from the grid resolution used to build the emulator; too low a resolution tends to result in biased estimates, whereas too high a resolution tends to result in biased estimates as well as a decrease in the computational efficiency. However as discussed above, the grid resolution effect seemed to be much smaller on predictive capabilities, implying that we are capturing the underlying spatio-temporal disease dynamics fairly well when using an emulator wherein the span of the design matrix successfully captures the true parameter values. As prediction is often going to be a primary goal, this would appear to be a promising result.

Emulator performance clearly depends upon the summary statistics of the epidemic data used to build the emulator; a problem that is common to both the classification methods of Nsoesie et al. (2011) or Pokharel & Deardon (2014), and so-called ABC techniques (McKinley, Cook, & Deardon, 2009; Numminen et al., 2013). Ideally a set of sufficient statistics will be available, but this is almost never the case. In contrast with the results of Pokharel & Deardon (2014), epidemic curve spatial stratification did not noticeably improve performance.

The use of more parsimonious summary statistics has also been considered by the authors; for example, various combinations of the peak of the epidemic curve, the time of last infection, and the proportion of individuals infected. It was found that such summary statistics produced emulators that performed less well than those based upon the entire epidemic curve. For example, when a standardized combination of the above three statistics was used to form the distance measure and applied to the simulation study of Section 3.1, for six out of ten simulated data sets analyzed both parameters were not captured in the respective marginal 95% PIs, in two analyses one parameter was not captured, and in the remaining two the other was not captured. Predictive performance was also less impressive with these summary statistics, although the deterioration in performance was not as marked as that regarding parameter estimation.

There are many other avenues open for further work. We here dealt with simulated data, or data that have previously been analyzed, providing prior information that would not be likely available, for example, in a real-life emergency outbreak. More work thus needs to be done to develop quick and easy methods to build the emulators—for example, adaptive methods in which design matrices of relatively low-resolution, but widely dispersed design points, could be used to do an initial analysis that provides results to redefine the design matrix for subsequent iterations.

We here chose relatively small data sets in which a full Bayesian MCMC analysis was computationally feasible. Of course when such an analysis is possible it makes far more sense to use a full, rather than approximate, Bayesian analysis. In reality emulation-based methods would only be used when such an analysis would not be computationally feasible. The corollary of this is, however, that in real life use there would be no gold-standard Bayesian analysis with which to compare and judge emulator-based inference and tune, for example, the design matrix. We have shown that in the disease systems considered here, however, epidemic curve prediction is reasonably good, even when parameter estimates are less impressive. Although good news in itself, it would be very useful to more exhaustively examine the practical effects of using the methods suggested here when, for example, using a fitted model to optimize a disease control strategy (e.g., Tildesley et al., 2006).

Finally our analysis was for a low-dimensional problem for a simple spatial disease system, here restricted to two model parameters. However when modelling such disease systems there are often things such as regression coefficients, seasonality parameters, and I period parameters that also require inferring. There would obviously be a large increase in the computational time required to simulate epidemics from each of the parameter combinations defined in the design matrix as the number of model parameters increases. Other authors have used GP emulators in

a similar context in slightly higher dimensional problems; for example, Jandarov et al. (2014) successfully fitted a gravity model with four parameters using a 20^4 grid resolution.

One way to alleviate the huge increase in computational burden as the number of model parameters increases would be to make use of parallel computation. Another would be simply to reduce the resolution of the grid defining the design matrix—although this would likely be at the cost of inferential performance. Of course there is no reason why a grid needs to be used to define the design matrix, we could make use of some sort of optimal design theory to define our design matrix in a high-dimensional space, setting the number of points we observe according to our computational resources. Emulators have been used in such situations to optimize high-dimensional stochastic models, albeit not in the context of inference; for example, Boukouvalas, Cornford, & Singer (2009) used a space filling design to build and then optimize an emulator for 14 input parameters. How emulator-based inference with more than a handful of parameters should be approached in practice is an area requiring further work.

It would often be unreasonable to assume that infection times and I periods are known. Thus we may want to include unknown latent period and I period parameters (leading to an $SEIR$ model). Additionally disease spread may propagate through an unknown (or only partially known) contact network that we wish to model. One advantage of the emulation approach to inference is that, although global parameters defining such things as the I period distribution and/or the contact network need to be inferred, nuisance parameters such as infection times, I periods, and/or the contact network itself, do not need to be explicitly inferred. They simply become part of the epidemic simulation. Preliminary results from work carried out by the authors imply that uncertain infection times and I periods in the SIR context can be easily and successfully incorporated into the analysis. However emulation-based methods do need to be tested and developed that can handle more complex systems, including the determination of suitable summary statistics for such systems.

ACKNOWLEDGEMENTS

This work was funded by the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA), the Natural Sciences and Engineering Research Council of Canada (NSERC), and was carried out on equipment funded by the Canada Foundation for Innovation—Leading Edge Fund project "Centre for Public Health and Zoonoses" at the University of Guelph.

BIBLIOGRAPHY

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2), 174–188.

Bastos, L. S. & O'Hagan, A. (2009). Diagnostics for Gaussian process emulators. *Technometrics*, 51(4), 425–438.

Bayarri, M. J., Berger, J. O., Cafeo, J., Garcia-Donato, G., Liu, F., Palomo, J., Parthasarathy, R. J., Paulo, R., Sacks, J., & Walsh, D. (2007). Computer model validation with functional output. *The Annals of Statistics*, 35, 1874–1906.

Beaumont, M., Cornuet, J.-M., Marin, J.-M., & Robert, C. P. (2009). Adaptive approximate Bayesian computation. *Biometrika*, 96(4), 983–990.

Bhat, K. S., Haran, M., & Goes, M. (2010). Computer model calibration with multivariate output: A case study. In *Frontiers of Statistical Decision Making and Bayesian Analysis*, Chen, M.-H., Dey, D. K., Müller, P., Sun, D., and Ye, K., editors, 168–184. Springer, New York.

Boukouvalas, A., Cornford, D., & Singer, A. (2009). Managing uncertainty in complex stochastic models: Design and emulation of a rabies model. *6th St. Petersburg Workshop on Simulation*, 839–841.

Brown, S., Csinos, A., Daíz-Pérez, J. C., Gitaitis, R., LaHue, S. S., Lewis, J., Martinez, N., McPherson, R., Mullis, S., Nischwitz, C., Riley, D., Sherwood, J., Sisson, V., Stephenson, M. G., & Wells, L. (2005). Tospoviruses in Solanaceae and other crops in the coastal plain of Georgia. Research Report 704, College of Agriculture and Environmental Science, University of Georgia.

Chis-Ster, I. & Ferguson, N. M. (2007). Transmission parameters of the 2001 foot and mouth epidemic in Great Britain. *PLoS ONE*, 2(6), e502.

Conti, S. & O'Hagan, A. (2010). Bayesian emulation of complex multi-output and dynamic computer models. *Journal of Statistical Planning and Inference*, 140, 640–651.

Cook, A. R., Gibson, G. J., Gottwald, T. R., & Gilligan, C. A. (2008). Constructing the effect of alternative intervention strategies on historic epidemics. *Journal of the Royal Society: Interface*, 5, 1203–1213.

Craig, P. S., Goldstein, M., Rougier, J. C., & Seheult, H. A. (2001). Bayesian forecasting for complex systems using computer simulators. *Journal of the American Statistical Association*, 96, 717–729.

Currin, C., Mitchell, T., Morris, M., & Ylvisaker, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, 86, 953–963.

Dancik, G. M. & Dorman, K. S. (2008). mlegp: Statistical analysis for computer models of biological systems using R. *Bioinformatics*, 24, 1966–1967.

Davis, R. A. (2001). Gaussian processes. *Encyclopedia of Environmetrics, Section on Stochastic Modeling and Environmental Change*, Brillinger, D., editor, Wiley, New York.

Deardon, R., Brooks, S. P., Grenfell, B. T., Keeling, M. J., Tildesley, M. J., Savill, N. J., Shaw, D. J., & Woolhouse, M. E. (2010). Inference for individual-level models of infectious diseases in large populations. *Statistica Sinica*, 20, 239–261.

Gamerman, D. & Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference.* Chapman & Hall/CRC, New York.

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

Haydon, D. T., Chase-Topping, M., Shaw, D. J., Matthews, L., Friar, J. K., Wilesmith, J., & Woolhouse, M. E. J. (2003). The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. *Proceedings of the Royal Society of London Series B - Biological Sciences*, 270, 121–127.

Hughes, G., McRoberts, N., Madden, L. V., & Nelson, S. C. (1997). Validating mathematical models of plant-disease progress in space and time. *IMA Journal of Mathematics Applied in Medicine and Biology*, 14, 85–112.

Jandarov, R., Haran, M., Bjørnstad, O., & Grenfell, B. (2014). Emulating a gravity model to infer the spatiotemporal dynamics of an infectious disease. *Journal of Royal Statitical Society, Series C*, 63(3), 423–444.

Kennedy, M. C. & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of Royal Statistical Society: Series B (Statistical Methodology)*, 63, 425–464.

King, A. A., Nguyen, D., & Ionides, E. L. (2015). Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, 69(12), 1–43.

Kwong, G. P. S. & Deardon, R. (2012). Linearized forms of individual-level models for large-scale spatial infectious disease systems. *Bulletin of Mathematical Biology*, 74(8), 1912–1937.

Liu, F. & West, M. (2004). A dynamic modelling strategy for Bayesian computer model emulation. *Bayesian Analysis*, 4(2), 393–412.

Marjoram, P., Molitor, J., Plagnol, V., & Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 15324–15328.

McKinley, T., Cook, A. R., & Deardon, R. (2009). Inference in epidemic models without likelhoods. *International Journal of Biostatistics*, 5(1).

Nsoesie, E. O., Beckman, R., Marathe, M., & Lewis, B. (2011). Prediction of an epidemic curve: A supervised classification approach. *Statistical Communication in Infectious Diseases*, 3(1):5.

Numminen, E., Cheng, L., Gyllenberg, M., & Corander, J. (2013). Estimating the transmission dynamics of streptococcus pneumoniae from strain prevalence data. *Biometrics*, 63(3), 748–757.

Oakley, J. E. & O'Hagan, A. (2002). Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, 89(4), 769–784.

O'Hagan, A. & Kingman, J. F. C. (1978). Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society, Series B*, 40(1), 1–42.

O'Neill, P. D. & Roberts, G. O. (1999). Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162, 121–129.

Pokharel, G. & Deardon, R. (2014). Supervised learning and prediction of spatial epidemics. *Spatial and Spatio-temporal Epidemiology*, 11, 59–77.

Sacks, J., Welch, W. J., Mitchell, T. J., & Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, 4(4), 409–423.

Santner, T. J., William, B. J., & Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer, New York.

Tildesley, M. J., Savill, N. J., Shaw, D. J., Deardon, R., Brooks, S. P., Woolhouse, M. E. J., Grenfell, B. T., & Keeling, M. J. (2006). Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. *Nature*, 440, 83–86.

Received 20 July 2015
Accepted 24 August 2016