Evidence-based controls for epidemics using spatio-temporal stochastic models in a Bayesian framework

Hola K. Adrakey[1, 2], George Streftaris[2], Nik J. Cunniffe[1], Tim R. Gottwald[3], Christopher A. Gilligan[1], and Gavin J. Gibson[2]

[1]Department of Plant Sciences, University of Cambridge, Cambridge, United Kingdom, CB2 3EA
[2]Maxwell Institute for Mathematical Sciences, School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, United Kingdom, EH14 4AS
[3]USDA Agricultural Research Service, 2001 South Rock Road, Fort Pierce, FL 34945, USA

Abstract

The control of highly infectious diseases of agricultural and plantation crops and livestock represents a key challenge in epidemiological and ecological modelling, with implemented control strategies often being controversial. Mathematical models, including the spatio-temporal stochastic models considered here, are playing an increasing role in the design of control as agencies seek to strengthen the evidence on which selected strategies are based. Here, we investigate a general approach to informing the choice of control strategies using spatio-temporal models within the Bayesian framework. We illustrate the approach for the case of strategies based on pre-emptive removal of individual hosts. For an exemplar model, using simulated data and historic data on an epidemic of Asiatic citrus canker in Florida, we assess a range of measures for prioritising individuals for removal that take account of observations of an emerging epidemic. These measures are based respectively on the potential infection hazard a host poses to susceptible individuals (hazard), the likelihood of infection of a host (risk) and a measure that combines both the hazard and risk (threat). We find that the threat measure typically leads to the most effective control strategies particularly for clustered epidemics when resources are scarce. The extension of the methods to a range of other settings is discussed. A key feature of the approach is the use of functional-model representations of the epidemic model to couple epidemic trajectories under different control strategies. This induces strong positive correlations between the epidemic outcomes under the respective controls, serving to reduce both the variance of the difference in outcomes and, consequently, the need for extensive simulation.

Keywords: Emerging epidemic | Spatio-temporal model | Non-centered parameterization | Control strategies| Bayesian Inference

1 Introduction

Highly infectious diseases of plants and arboreal populations such as Asiatic citrus canker, Huanglongbing, ash dieback, sudden oak death, or veterinary pathogens such as foot-and-mouth disease and classical swine fever represent a major threat at both the global and the regional level and lead to significant economical losses (Ferguson et al., 2001; Schubert et al., 2001; Gottwald et al., 2001, 2002b; Filipe et al., 2012; DEFRA, 2013; Thompson et al., 2004; Cunniffe et al., 2016; Thompson et al., 2016). Considerable resources are deployed to control the spread of these and other diseases (Schubert et al., 2001; USDA/APHIS et al., 2006; Parnell et al., 2009; DEFRA, 2013; Cunniffe et al., 2014). An approach commonly adopted to control a disease outbreak is to remove susceptible individuals from a population, for example from a neighbourhood of a detected infectious host. Controls of this kind have frequently proved controversial on account of their socio-economic and other impacts on farmers or other stakeholders that they affect (Schubert et al., 2001; Graham et al., 2004; Ferguson et al., 2001; Gottwald et al., 2002b). An important challenge, therefore, is that of optimising control strategies so that they provide the greatest benefits in terms of disease reduction for a given level of control (Cunniffe et al., 2015).

We address this challenge in the context of an epidemic of an infectious disease that spreads through a population of spatially distributed hosts, and is controlled by testing and removing individual hosts (if found to be infected), *via* the objectives of:

(i) presenting a computational statistical framework within which competing control strategies for an emerging epidemic can be represented and their likely efficacy assessed in the light of available data in a computationally efficient manner;

(ii) illustrating the use of the framework in a particular scenario – a spatio-temporal epidemic driven by SI dynamics and controlled by removal of hosts – to formulate and to explore the relative merits of competing strategies for selecting hosts for removal;

(iii) describing how the framework can be applied to design controls for alternative choices of epidemic model or control mechanisms.

In order to develop the framework and illustrate its use we consider epidemics for which infection can be spatially-dependent so that the infectious challenge presented to a susceptible host by a given infected individual is dependent on the distance between them. This leads us to consider epidemics that can be represented using individual-based, spatio-temporal stochastic models. The 'individual' in such formulations may represent an individual host or a larger conglomeration of hosts such as a field, farm, plantation or a village, making the general class of models we consider very flexible in terms of the host-pathogen systems to which it is relevant. We assume that partial observations on an emerging epidemic are available to inform the actions that are taken at some specified future time to control subsequent spread. We consider explicitly only controls that involve the removal of infected or susceptible individuals from the population. Throughout we will assume that constraints are placed on the level of resource that can be expended on a control strategy. These could take the form of bounds on the numbers of individuals that can be removed, the spatial area that can be surveyed, or the number of separate regions to which control can be applied. The problem is then to identify the optimal control strategy satisfying these constraints.

To achieve a coherent approach for the model-based design of an efficient control that allocates available resources to maximise the impact on the spread of the epidemic, we work within the Bayesian framework. As explained in Section 2 we use posterior predictive expectations of certain quantities associated with a developing epidemic both to assess the effectiveness of controls, and to prioritise those individuals or regions that should be targeted using a control strategy. In particular, we will investigate several approaches to constructing a geographical map prioritising sites or regions according to a range of candidate measures. Similar ideas have been used in Boender et al. (2007), te Beest et al. (2011) and Hyatt-Twynam et al. (2017) where the map is constructed on the basis of combining the basic reproduction number with estimates of the probability of infection. A key feature of the approach in this paper is the use of non-centred parameterisations of epidemic models (specifically based on the Sellke construction (Sellke, 1983)) in order to couple the trajectories of epidemics simulated from their respective posterior predictive distributions under different control strategies. This idea has already been applied by some of the authors (Cook et al., 2008) for retrospective assessment of controls. In this paper we apply it in the context of prospective control where the task is to select control strategies to impact on the future trajectory of an epidemic in progress. As proposed in Section 2, and demonstrated in Section 3 the approach has the potential to reduce the amount of simulation required to estimate the expected differences in effectiveness of different control strategies - essentially by reducing the variance of these differences. Using this approach we are able to dispense with the need to nest extensive simulation within optimisation algorithms in delivering computationally efficient schemes.

Although the methods may be developed for a specific scenario they are designed to be generally applicable across a range of scenarios. Therefore, in keeping with objective (iii) above, in Section 4 we present in outline how the methods can be adapted to epidemic models with more complex interactions that are controlled by different strategies, or observed with imperfect diagnostics.

The paper is organised as follows. In Section 2 we introduce the class of model processes and outline the Bayesian computational approaches that we use. We also describe how we can exploit non-centered parameterisations in order to couple stochastic epidemics under competing control strategies and to reduce the variance of comparative performance estimators. We present the quantitative measures whose posterior predictive expectations will be used to prioritise the application of control. Section 3 illustrates the application of the methods to optimise control strategies in simulated and real-world scenarios. Conclusions, potential extension of the methods and avenues for further research are discussed in Section 4.

2 Materials and Methods

2.1 Epidemiological models

We consider a spatially-explicit, stochastic, individual-based, compartmental SI model (Neri et al., 2014) for the spread of an infectious disease through a discrete population in a bounded region. Hosts are identified by their location vectors which may take values in a continuous space or, as in the case of a managed arboreal population, may lie on the vertices of a rectangular lattice. At any time t, hosts can be partitioned into two classes $S(t)$ and $I(t)$, containing susceptible and infected individuals respectively. We further assume that $I(t)$ can be partitioned into two groups $I_c(t)$ and $I_s(t)$ denoting cryptic and symptomatic infections respectively. It is assumed that individuals in $I_s(t)$ are obviously infected but those in $I_c(t)$ can only be determined using some diagnostic test. Suppose that i represents a susceptible host at time t. Then the probability that i is infected in the period $[t, t+dt]$ is given by the following equation:

$$P\left(i \text{ infected in } [t, t+dt]\right) = \lambda_i(t)dt + o(dt), \tag{1}$$

where

$$\lambda_i(t) = \left(\beta \sum_{j \in I(t)} K(d_{ji}, \alpha) + \epsilon\right) \tag{2}$$

is the force of infection on host i at time t, β is the contact parameter and ϵ the primary infection rate, this being the rate at which any individual i contracts the disease from an external or environmental source. In addition, $K(d_{ji}, \alpha)$ is a non-negative function characterizing the infection challenge posed by the host j to i as a function of the inter-host distance d_{ji}, and known as the dispersal kernel with parameter α (the dispersal parameter). In typical formulations, for any given α, the function K decreases with the distance. Intuitively, the instantaneous rate at which i is becoming infected, $\lambda_i(t)$, is composed of the sum of the infection rate from environmental sources and the individual infection rates from infected individuals at time t.

Moreover, we assume for simplicity that, following infection, individuals remain asymptomatic (i.e. in $I_c(t)$) for a fixed, known period of time Δ, before moving to $I_s(t)$. In more general formulations, the sojourn time in the cryptic compartment could be modelled by assigning an appropriate distribution, for example a Gamma or Weibull distribution (Parry et al., 2014). The fact that asymptomatic hosts are only identifiable through some diagnostic test presents challenges for the design of controls as both symptomatic and cryptic infections present a threat to susceptible individuals in the population. The model described above has been successfully applied to plant diseases, including diseases of citrus such as Asiatic citrus canker, where disease-induced mortality occurs at a far longer timescale than epidemic spread and control intervention. With some modification it can be applied to natural plant populations or to veterinary epidemics spreading through populations of farms (Tildesley et al., 2006; Jewell et al., 2009) where the infectivity of farms may vary with the particular species mix. The definition of realistic distance measures for populations of farms is challenging since the connectivity between pairs of farms is affected by factors such as animal movements to and from market places as well as Euclidean distance. Additional compartments - such as an exposed class E, in which hosts are infected but not yet able to infect, or a removed class R, representing host removal by death, acquisition of immunity, or other means can be included. Note that for the basic SI model considered here, in the absence of control, the number of infected individuals in the population would increase monotonically until the entire population were infected.

2.2 Sellke construction

Following the idea developed in Sellke (1983), we consider each susceptible host j to possess a level of resistance to the infection pressure quantified by a threshold Q_j, known as the Sellke threshold, where $Q_j \sim Exp(1)$, and thresholds are independent over hosts. During the epidemic process, the cumulative pressure on an individual j by time t is given by the integral $A_j(t) = \int_0^t \lambda_j(u)du$. Individual j becomes infected at the time t_j for which $Q_j = A_j(t)$, this being the time at which the accumulated infectious pressure reaches the threshold Q_j. This description is equivalent to the standard stochastic process given by the equation (1).

Now, given the parameter $\boldsymbol{\theta} = (\alpha, \beta, \epsilon)$ and given the set of Sellke thresholds $\underline{Q} = (Q_1, Q_2, \ldots, Q_N)$ the trajectory is uniquely specified in the absence of control. Moreover, for a control $\boldsymbol{d}$ that involves surveying, testing and removing infected hosts at particular times, then (assuming a perfect test for detecting infection) the epidemic trajectory is uniquely specified by $(\boldsymbol{\theta}, \underline{Q}, \boldsymbol{d})$. The particular benefit from using this representation in the context of this paper derives from the fact that a combination of parameters and threshold $(\boldsymbol{\theta}, \underline{Q})$ of thresholds uniquely specifies the epidemic outcome that arises for any control strategy based on removal of hosts. This will be particularly useful

when we wish to compare the effect of two interventions on the same set of hosts; more precisely we can couple epidemics under different control strategies by merely matching latent processes (Cook et al., 2008).

2.3 Observation process and control problem

We consider the following situation (see Figure 1). We assume that observations on an emerging epidemic are collected over a period of time $[t_0, t_{obs}]$ with no control applied during this period. We denote by $\mathbf{y}$ the data observed up to and including t_{obs} which may consist of a sequence of 'snapshots' of the symptomatic set of hosts at discrete times, or other forms of partial data. We assume that the epidemic proceeds according to the model of Section 2.1 with unknown parameter vector $\boldsymbol{\theta}$. We define the *trajectory* of the epidemic up to any time t to be $\underline{x}(t)$ where $\underline{x}(t)$ specifies the time and nature of every transition occurring during $[t_0, t]$. The intervention (control) time when the control is applied is denoted by $t_C > t_{obs}$ and we denote by $t_A \geq t_C$ the assessment time at which the effectiveness of the control is quantified (e.g. in terms of the numbers of infections up to t_A). We define an *impact function* $u(\underline{x}(t))$ in order to quantify the practical significance of an epidemic with the purpose of control being to minimise this function. Although alternatives could be selected, throughout this paper we define $u(\underline{x}(t))$ to be the total number of hosts infected by time t. Therefore the effectiveness of any control will be determined from consideration of $\underline{x}(t_A)$.

Let $\pi(\boldsymbol{\theta})$ denote a prior density for the model parameter vector which represents our belief about $\boldsymbol{\theta}$ at time t_0. We denote by $\pi_0(\underline{x}(t)|\mathbf{y})$ and $\pi_{\boldsymbol{d}}(\underline{x}(t)|\mathbf{y})$ the posterior distribution, given $\mathbf{y}$, of the trajectory of the epidemic up to time t subject to no control and control $\boldsymbol{d}$ respectively. For any control $\boldsymbol{d}$ and assessment time t_A, we define the expected impact conditional on the observed data, $\mathbf{y}$, to be

$$U(\boldsymbol{d}, t_A) = E_{\boldsymbol{d}}\left(u(\underline{x}(t_A))|\mathbf{y}\right) = \int u(\underline{x}'(t_A))\pi_{\boldsymbol{d}}\left(\underline{x}'(t_A)|\mathbf{y}\right) d\underline{x}'(t_A). \tag{3}$$

We define the optimal control as that which minimises $U(\boldsymbol{d}, t_A)$.

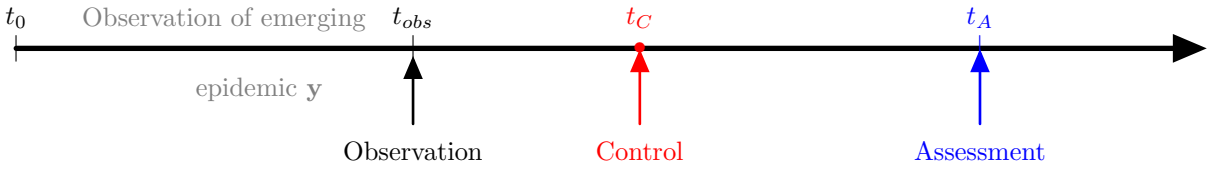

Figure 1: Graphical representation of the observation-control-impact system. Given observations of the system from some initial time t_0 up to t_{obs} a subset of hosts is considered for potential removal at time t_C (if infected at t_C). The impact of the control strategy is assessed at assessment time t_A by considering the history of the epidemic up to t_A.

2.4 Comparing control strategies

A straightforward approach to simulation-based optimal design for this scenario is to utilise Monte Carlo simulation by drawing samples $(\underline{x}(t_A), \boldsymbol{\theta})$ from $\pi_d(\underline{x}(t_A), \boldsymbol{\theta}|\mathbf{y})$ to generate a sample from $\pi_{\boldsymbol{d}}(\underline{x}(t_A)|\mathbf{y})$ from which $U(\boldsymbol{d}, t_A)$ can be estimated, and carrying this out independently over different controls $\boldsymbol{d}$. This in essence is the approach taken by Cunniffe et al. (2015) where controls are compared on simulated replicates using the Gillespie algorithm (Gillespie, 1977), although without estimating model parameters. Here we use the Sellke construction to give a more efficient sampling strategy. We exploit the fact that the epidemic trajectory is uniquely specified by $(\theta, \underline{Q}, \boldsymbol{d})$ so that

$$\underline{x}(t) = h(\theta, \underline{Q}, \boldsymbol{d}, t) \tag{4}$$

for any t. Specifically we draw a random sample $\{(\theta_i, \underline{Q}_i)|i = 1, \ldots, m\}$ from $\pi(\boldsymbol{\theta}, \underline{Q}|\mathbf{y})$. Then, for any control $\boldsymbol{d}$, we can obtain a random sample from $\pi_{\boldsymbol{d}}(\boldsymbol{\theta}, \underline{x}(t_A)|\mathbf{y})$ as $\{h(\theta_i, \underline{Q}_i, \boldsymbol{d}, t_A)|i = 1, \ldots, m\}$, using the algorithm described in Section 2 of the electronic supplementary material (ESM). The coupling of trajectories under different controls $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$ but with common $(\theta, \underline{Q})$ should ideally induce a strong positive correlation between the numbers of infected hosts associated with the control scenarios $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$, $u(h(\theta, \underline{Q}, \boldsymbol{d}_1, t_A))$ and $u(h(\theta, \underline{Q}, \boldsymbol{d}_2, t_A))$, bringing benefits in reducing the variance of $u(h(\theta, \underline{Q}, \boldsymbol{d}_1, t_A)) - u(h(\theta, \underline{Q}, \boldsymbol{d}_2, t_A))$ and, hence, the variance of $\hat{U}(\boldsymbol{d}_1, t_A) - \hat{U}(\boldsymbol{d}_2, t_A)$ where

$$\hat{U}(\boldsymbol{d}, t_A) = \frac{1}{m}\sum_{i=1}^{m} u(h(\theta_i, \underline{Q}_i, \boldsymbol{d}, t_A))$$

2.5 Removal-based control strategies

We mainly consider control measures based on the removal of hosts in which infection is detected. While symptomatic hosts are visually detectable, we assume that although a host, cryptic at the time of a survey contributing to $\mathbf{y}$, will not be recorded as infected in that survey, any infection is observable during the control phase, thanks to the availability of a diagnostic test.

We assume that control in the form of removal of hosts is to be implemented at time t_C and assume that the availability of resources dictates that only N' hosts can be considered for potential removal. Any host that is found to be infected (either because it shows visible symptoms or because a diagnostic test reveals that it is cryptically infected) is removed. However any host that is not infected remains in the population. We note that, for simplicity, the diagnostic tests considered here are assumed to have perfect sensitivity and specificity. This is rarely the case in practice and we later discuss how this assumption may be relaxed. While this paper focuses on this particular form of control, the general methods could be applied to design controls based on alternative strategies such as ring culling. Our aim here is to compare strategies for prioritising the N' hosts considered for control (removal of infection detected) in terms of their respective expected impact on the epidemic size.

2.6 Prioritisation scheme

We now describe the measures used as criteria for host prioritisation. For each host, we construct a range of metrics subsequently used to prioritise hosts for consideration under a given control strategy.

The measures used can all be expressed as $E(G^j(\underline{x}(t_M))|\mathbf{y})$, the posterior expectation of some function of the system state at some time $t_M \geq t_{obs}$ for host j, under the assumption that no control is deployed. This general concept has been previously used in the literature to target priority sites (Boender et al., 2007; Tildesley et al., 2009; Kao, 2003; te Beest et al., 2011; DEFRA, 2013; Cunniffe et al., 2015; Hyatt-Twynam et al., 2017). Typically, the candidate hosts with the highest measure are prioritised.

Here, for any t_M, for each host we let $G_R^j(\underline{x}(t_M))$ and $G_H^j(\underline{x}(t_M))$ respectively denote the infection status of j at t_M under trajectory $\underline{x}(t_M)$ and the infectious challenge posed to the remaining susceptibles if that host were infected at time t_M. More formally, the *risk* measure is given by

$$\mathcal{R}_j(t_M) = E\left(G_R^j(\underline{x}(t_M))|\mathbf{y}\right) \tag{5}$$

where

$$G_R^j(\underline{x}(t_M)) = \mathbb{1}_{\{x_j \leq t_M\}}, \tag{6}$$

x_j is the infection time of host j and $\mathbb{1}$ is the indicator function. Hence the risk measure, evaluated at t_M, for a given host simply represents the posterior probability that the host is infected at time t_M. The *hazard* is defined as

$$\mathcal{H}_j(t_M) = E\left(G_H^j(\underline{x}(t_M))|\mathbf{y}\right) \tag{7}$$

where

$$G_H^j(\underline{x}(t_M)) = \beta\sum_{i\neq j} K(d_{ij}, \alpha)\mathbb{1}_{\{x_i > t_M\}} \tag{8}$$

The hazard measure is designed to quantify how much infectious challenge a given host could present at time t_M taking account of where it is located with respect to the remaining susceptible population at that time.

In DEFRA (2013), it has been argued that considering such measures in isolation for prioritisation may not be cost-effective. For example removing a host with high risk might be less cost-effective if it is unlikely to infect other hosts in the population. It was concluded that a measure that combines the likelihood of infection with the propensity to infect susceptibles will provide the best prioritisation scheme (DEFRA, 2013). Developing this idea, we define a further measure to represent the *threat* posed by each host at time t given the observed data $\mathbf{y}$ as

$$\mathcal{T}_j(t_M) = E\left(G_T^j(\underline{x}(t_M))|\mathbf{y}\right) \tag{9}$$

where

$$G_T^j\left(\underline{x}(t_M)\right) = G_R^j(\underline{x}(t_M))G_H^j\left(\underline{x}(t_M)\right) \tag{10}$$

The threat measure therefore represents the posterior expectation of the infectious challenge presented by any given host j to susceptibles at time t_M and, consequently, represents the expected reduction in infectious challenge that would result from consideration of this host in the control strategy.

2.7 Data and Inference

We suppose that the data $\mathbf{y}$ consist of a sequence of snapshots observed at particular times in $[t_0, t_{obs}]$. As prioritisation and assessment measures require prediction of the trajectory of the epidemic at times beyond t_{obs} they are best treated using Bayesian data-augmentation approaches (Neri et al., 2014; Parry et al., 2014; Lau et al., 2015). We use a noninformative prior $\pi(\boldsymbol{\theta})$ for the model parameter vector by assigning independent, vague uniform priors to α, β and ϵ. We then 'augment' $\boldsymbol{\theta}$ with the unobserved epidemic trajectory $\underline{x}(T)$, where $T \geq t_{obs}$ and use Markov chain Monte Carlo (MCMC) to draw samples from the joint posterior density $\pi(\boldsymbol{\theta}, \underline{x}(T)|\mathbf{y}) \propto \pi(\theta)\pi(\underline{x}(T)|\boldsymbol{\theta})\Pr(\mathbf{y}|\underline{x}(T))$, this being a standard approach in fitting stochastic spatio-temporal models. Note that, for the 'snapshot' observational model assumed here, the term $\Pr(\mathbf{y}|\underline{x}(T))$ is 0 or 1 depending on whether $\underline{x}(T)$ would yield the data $\mathbf{y}$.

All inferences carried out from here on are based on an investigation of the posterior density $\pi_0(\boldsymbol{\theta}, \underline{x}(T)|\mathbf{y})$ where T can be chosen in a number of ways. First note that the data $\mathbf{y}$, being a sequence of snapshots of symptomatic sets of hosts can be interpreted as specifying a period for the infection of each symptomatic host of the form $[\tau_{j-1} - \Delta, \tau_j - \Delta]$ where τ_j is the time at which the host was first observed as symptomatic and Δ is the cryptic period defined in Section 2.1. It follows that a suitable algorithm could be designed by setting $T = t_{obs} - \Delta$, as the data in effect distinguish hosts infected before $t_{obs} - \Delta$ from those infected after $t_{obs} - \Delta$. However, given the need to impute infections beyond $t_{obs} - \Delta$ to investigate the posterior distribution of the prioritisation measures at t_M, we implement a more general algorithm with $T > t_{obs} - \Delta$. This is done using methods which are now standard in computational epidemiology. Details of algorithms are given in Section 1 of the ESM.

2.8 Calculation of prioritisation measures and imputation of Sellke thresholds

The calculation of the risk, hazard and threat measures is achieved by imputing the functions $G^j\left(\underline{x}(t_M)\right)$ using the imputed $(\boldsymbol{\theta}, \underline{x}(t_M))$ and is straightforward using equations (5), (7) and (10). For each draw $(\boldsymbol{\theta}^{(k)}, \underline{x}(t_M)^k) \sim \pi_0(\boldsymbol{\theta}, \underline{x}(t_M)|\mathbf{y})$, the vectors

$$\underline{G}_R^{(k)}(\underline{x}(t_M)) = (G_R^{1(k)}(\underline{x}(t_M)), \ldots, G_R^{N(k)}(\underline{x}(t_M)))$$

and

$$\underline{G}_H^{(k)}(\underline{x}(t_M)) = (G_H^{1(k)}(\underline{x}(t_M)), \ldots, G_H^{N(k)}(\underline{x}(t_M)))$$

are computed to provide a sample from the joint posterior distribution $\pi_0(\underline{G}_R^j(\underline{x}(t_M)), \underline{G}_H^j(\underline{x}(t_M))|\mathbf{y})$ for $1, \ldots, N$.

The risk, hazard and threat measures defined in equation (5), (7) and (10) are then approximated using the Monte Carlo approximation respectively by:

$$\mathcal{R}_j^*(t_M) = \frac{1}{m}\sum_{k=1}^{m} G_R^{j(k)}\left(\underline{x}(t_M)\right) \tag{11}$$

$$\mathcal{H}_j^*(t_M) = \frac{1}{m}\sum_{k=1}^{m} G_C^{j(k)}\left(\underline{x}(t_M)\right) \tag{12}$$

$$\mathcal{T}_j^*(t_M) = \frac{1}{m}\sum_{k=1}^{m} \left(G_R^{j(k)}\left(\underline{x}(t_M)\right) G_H^{j(k)}\left(\underline{x}(t_M)\right)\right) \tag{13}$$

where m is the number of draws generated from $\pi_0(\boldsymbol{\theta}, \underline{x}(t)|\mathbf{y})$.

As our approach to comparing the effectiveness of controls relies on coupling epidemics assuming common sets of Sellke thresholds, we impute the latter explicitly using samples from the MCMC algorithm. For any $T > t_{obs}$, given a draw $(\boldsymbol{\theta}, \underline{x}(T))$ from $\pi_0(\boldsymbol{\theta}, \underline{x}(T)|\mathbf{y})$ we can impute the Sellke thresholds Q as follows:

$$Q_j = \begin{cases} \int_0^{t_j} \left(\beta \sum_{i \in I(u)} K(d_{ij}, \alpha) + \epsilon\right) du & \text{if } j \text{ is infected at } t_j < T \\ \int_0^{T} \left(\beta \sum_{i \in I(u)} K(d_{ij}, \alpha) + \epsilon\right) du + \zeta & \text{if } j \text{ is susceptible at } T \end{cases} \tag{14}$$

where $\zeta \sim Exp(1)$. Given a random draw $(\boldsymbol{\theta}, \underline{x}(T)) \sim \pi_0(\boldsymbol{\theta}, \underline{x}(T)|\mathbf{y})$, it is straightforward to use the construction in the equation (14) to impute the corresponding Sellke thresholds $\boldsymbol{Q}$ and to convert a sample of points from $\pi_0(\boldsymbol{\theta}, \underline{x}(T)|\mathbf{y})$ to a sample from the joint posterior distribution of the parameter and the thresholds, $\pi_0(\boldsymbol{\theta}, \underline{Q}|\mathbf{y})$. A random sample from the posterior distribution $(\boldsymbol{\theta}, \boldsymbol{Q}) \sim \pi_0(\boldsymbol{\theta}, \boldsymbol{Q}|\mathbf{y})$ is used as a population of 'pre-epidemics' on which subsequent analyses to compare controls can be based. Once the population of 'pre-epidemics' has been generated, subsequent computations for assessing controls become entirely deterministic.

3 Applications to simulated and real-world host populations

3.1 Uniformly distributed host population

We test the methodology on a spatio-temporal epidemic simulated in a population of size $N = 1000$, with host locations sampled independently from a uniform distribution over a $0.75 \times 0.75\text{km}^2$ square region (Figure 2 of the ESM). The observations are made between $t_0 = 0$ (time corresponding to the introduction of the external source of infection) and $t_{obs} = 460$ and consist of a sequence of snapshots of a symptomatic set of hosts taken at 30−day intervals. The entire population is assumed susceptible at $t_0 = 0$ and the process is governed by the equation (1). We use $\alpha = 0.08\text{km}$, $\beta = 7 \times 10^{-6}\text{days}^{-1}\text{km}^2$ and $\epsilon = 5 \times 10^{-5}\text{days}^{-1}$ for the simulation and consider an exponential kernel $K(d, \alpha) = \frac{1}{2\pi d\alpha} \exp(-d/\alpha)$. The parameters along with the kernel reflect the findings in Neri et al. (2014). The choice of the primary infection rate ϵ ensures that if all hosts are susceptible, we expect one primary infection around every 20 days, reflecting the typical epidemic in Broward county (region B2 in Neri et al. (2014)) where the first infection was detected within the first month of the observation. Moreover, we set the time taken for symptoms to appear following an infection to be $\Delta = 100$ days, representing the assumptions used for Asiatic citrus canker by Neri et al. (2014). As discussed earlier, the data $\mathbf{y}$ effectively specify an interval for the infection time of each symptomatic host. At time t_{obs} there are 128 symptomatic hosts while 153 are undetected (cryptic) infections. The epidemic progress is shown in Figure 2 of the ESM.

We use the MCMC routines described in the ESM to sample from the posterior distribution $\pi_0(\boldsymbol{\theta}, \underline{x}(T)|\mathbf{y})$. Non-informative uniform priors $U[0, 1000]$ are used for all parameters. To validate the implementation of the methods we repeat the estimation for $T = t_{obs} - \Delta, T = t_{obs}$, and $T = t_A = 500$, the assessment time used later, noting that the marginal $\pi_0(\boldsymbol{\theta}|\mathbf{y})$ should be the same in all cases. Note that the last two cases require the use of reversible-jump methods as the number of infection events in $\underline{x}(T)$ is not fixed by the data. Details of the MCMC runs are found in Section 3 of the ESM. We note that the estimated densities are invariant over the values of T and that parameter values used for the simulation are consistent with their respective posterior densities. Note that the posterior distributions shown in Figure 3 of the ESM exhibit considerable uncertainty regarding the values of α, β and ϵ showing that these parameters cannot be estimated precisely from the observations available up to t_{obs}. Nevertheless, the Bayesian framework naturally allows us to take account of this parameter uncertainty when predicting the future trajectory of the epidemic and the impact of controls.

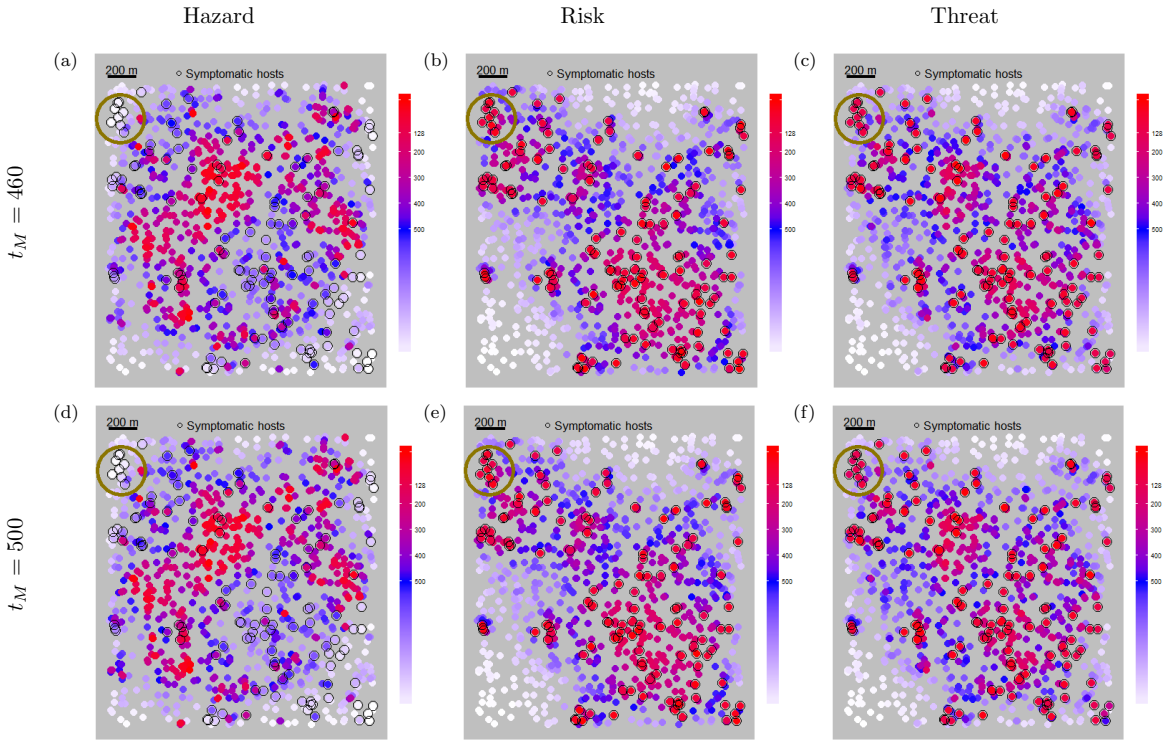

Figure 2: Posterior predictive maps of the hazard ((a) and (d)), risk ((b) and (e)) and threat ((c) and (f)) measures calculated for $t_M = t_{obs}$ and $t_M = t_A$ using equations (11-13) for the simulated epidemic on the uniformly distributed host population (Section 3.1). Each circle represents an individual host with colour varying from white to blue to red with increasing values of the respective measure for that host. The 128 symptomatic hosts detected during the survey are indicated by the black circles. Note that the hazard values ((a) and (d)) are greatest in regions of low infection while the risk measure is greatest for symptomatic individuals. The dependence of the threat measure on the positions of likely susceptible individuals in relation to an infected host can be discerned. For example, the infected hosts (circled) in the top left corner of the population naturally exhibit high values of the risk while the corresponding threat measure is comparatively lower for these hosts, as a high proportion of their immediate neighbours are already infected.

We now consider the effect on implementing alternative controls, as described in Section 2.5 at time $t_C = 460$, for this simulated epidemic using the three prioritisation schemes of Section 2.6, where measures are computed from $\pi_0(\underline{x}(t_M)|\mathbf{y})$ with $t_M = t_C$ and $t_M = t_A$. The resulting maps, which appear largely similar for $t_M = t_C$ and $t_M = t_A$, are displayed in Figure 2.

Controls are compared using the performance measures of Section 2.4. Figure 3 shows the estimated values of the expected number of infections and the estimated expected reduction (with respect to the uncontrolled scenario) respectively for the three prioritisation schemes based on risk, hazard and threat map respectively and how this varies with N', the number of hosts considered. Measures are estimated using a sample of size $m = 1000$ from $\pi(\theta, \underline{Q}|\boldsymbol{y})$. Note that the minimum value of N' is chosen to be 128, reflecting the case where the risk measure selects the 128 symptomatic sites for removal. For $N' < 128$ a further sampling scheme would be required to select the hosts to be considered under the risk measure $\mathcal{R}$.

Since, for any of the control strategies (accept that based on $\mathcal{R}$ with $N' = 128$), it is likely that fewer than N' hosts are removed, we can effect a further comparison of the prioritisation schemes on the basis of the expected number of hosts removed using each, estimated from the $m = 1000$ realisations of $(\boldsymbol{\theta}, \underline{Q})$. These are plotted against N' in Figure 3 for the 3 schemes. These results highlight the efficiency of the scheme based on $\mathcal{T}$ which achieves the best reduction in expected number of infections at the assessment time, t_A. On the other hand, Figure 3 shows that the controls designed using the risk and threat measures give similar performance, highlighted by their respective maps (see Figure 2). This phenomenon may conceivably arise due to the relatively homogeneous spatial structure of the host population and the resulting epidemic that is observed for the particular choice of parameters. As a result,

the imputed values of $G_H(\underline{x}(t))$ may not exhibit great variability over hosts, suggesting that the values of $G_R(\underline{x}(t))$ may have the greater influence in determining the threat map. This partly motivates our consideration in Section 3.2 of heterogeneously structured populations. We further note that there is little difference in the effectiveness of controls using prioritisation maps evaluated at $t_M = t_C$ and $t_M = t_A$, as may be predicted from the similarity of the maps in Figure 2.

In Figures 3(a) and 3(d) the confidence intervals for the mean number of infections by t_A appear quite wide, reflecting the large variance of the predictive distribution of the numbers of infections. By contrast, the confidence intervals for the mean reduction in comparison to the no-control case (see Figures 3(b) and 3(e)) are narrow. This contrast is due to the strong positive correlation that is induced between the numbers of infections by t_A under different control regimes when the respective epidemic trajectories are driven by the same set of Sellke thresholds and parameter values. This positive correlation then reduces the variance of the *difference* between the numbers of infections, narrowing the confidence interval for the mean difference.

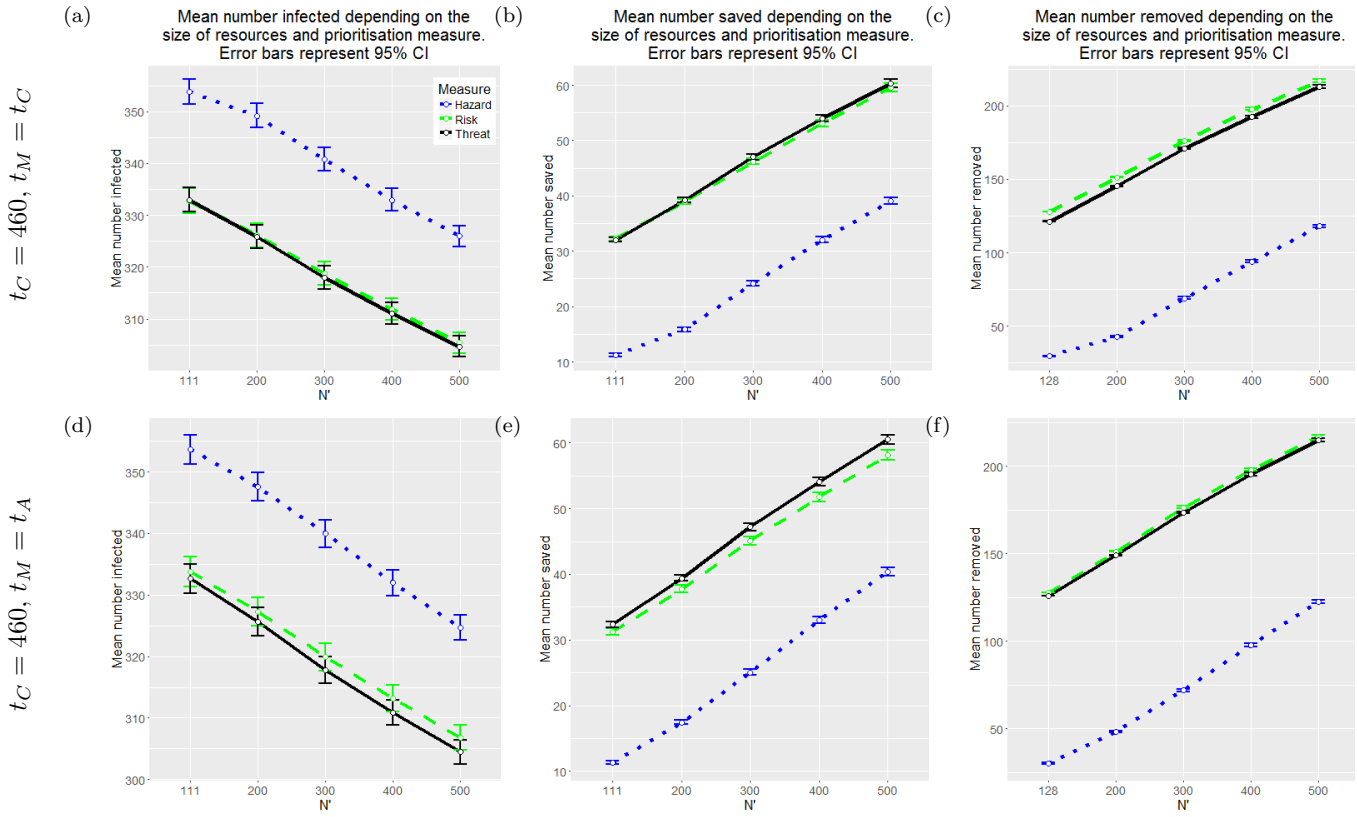

Figure 3: Marginal confidence intervals for the expected number of infections by t_A ((a) and (d)), the estimated expected reduction in infection with respect to the no-control case ((b) and (e)), and the expected number of removed hosts ((c) and (f)), when maps are constructed at t_{obs} ((a)-(c)) and t_A ((d)-(f)), for a range of values of N', the number of hosts considered for removal.

3.2 Application to structured populations: citrus locations from Florida

To illustrate the approach described above on a clustered host population, we use data regarding citrus locations from Florida to mimic a realistic spatial distribution of hosts, through which we consider the spread and control of an epidemic of Asiatic citrus canker, previously analysed by Neri et al. (2014).

3.2.1 Simulated data

The data used for the analysis consist of the citrus locations from a site located in Broward county, labelled B_2 from the four sites in an urban region close to Miami (Gottwald et al., 2002a,b; Neri et al., 2014). A total of $18,769$ trees across the four sites were monitored with $1,111$ in B_2.

The locations of the citrus population are then used to simulate epidemics governed by Equation (1). Two different epidemics are simulated using the normalised exponential kernel considered in Neri et al. (2014), with and without primary infection. The kernel takes the form

$$K(d,\alpha) = \frac{1}{2\pi d}\frac{1}{\alpha}\exp(-d/\alpha) \tag{15}$$

where d is the Euclidean distance between infected and susceptible hosts.

- Case(I): An exponential kernel with primary infection

We assume that the entire population is susceptible at time $t_0 = 0$, the time corresponding to the introduction of the external source. The value used for the contact rate, the dispersal parameter and the primary infection rate are respectively $\beta = 7 \times 10^{-6}\,\text{days}^{-1}\text{km}^2$, $\alpha = 0.08\,\text{km}$, $\epsilon = 5 \times 10^{-5}\,\text{days}^{-1}$ and we observe the process up to time $t_{obs} = 460$ days by which time 169 hosts were symptomatic with 133 cryptic. Figure 4 shows the progress of the simulation over time. The parameters are chosen from Neri et al. (2014) where they were estimated *via* MCMC using 12 months of the epidemiological data.

- Case(II): An exponential kernel with no primary infection.

We perform a similar experiment with $\beta = 8 \times 10^{-6}\text{days}^{-1}\text{km}^2$, $\alpha = 0.8\text{km}$ and $\epsilon = 0$ but assuming that $t = 0$ corresponds to the time of the initial infection. For convenience, we choose the first infection from the Canker data (Neri et al., 2014) to be the host initially infected. Here, we maintain $t_{obs} = 460$ and we observe 111 symptomatic and 124 cryptic individuals at this time (see Figure 5 for the progress of a simulation over time).

Although symptoms can be seen within $10 - 14$ days, the average time to symptom discovery in residential trees was 108 days (Gottwald et al., 2002b). Here we again use $\Delta = 100$ days post-infection as a convenience, in line with the assumption by Parnell et al. (2009) and Neri et al. (2014).

For parameter estimation, we again adopt the MCMC algorithm described in Section 1 of the ESM using vague $U[0, 1000]$ priors on the model parameters. The estimation is done as in Section 3.1 with T varying depending on the case considered. The posterior distributions of the model parameters α, β and ϵ for various T shown in Figure 6 of the ESM match, regardless of how far we impute infection times beyond t_{obs}. This provides some evidence that the algorithm gives an accurate picture of the posterior distribution.

3.2.2 Results

We show the effectiveness of controls developed using the three measures constructed in Section 2.6. We consider two possible times for the implementation of control, $t_C = 460$ and $t_C = 470$ and, for each value of t_C, we consider the cases respectively for $t_M = t_C$ and $t_M = t_A$. Again, these measures are computed by drawing 10^5 samples from $\pi_0(\underline{x}(t)|\mathbf{y})$ at $t = t_C$ and $t = t_A$. Figures 6 and 7 show the maps for the cases with and without primary infection respectively. We note some apparent differences between risk and threat maps with the latter having a tendency to prioritise sites around the periphery of the cluster of infected sites. We present in Figures 8 and 9 the effect of varying N' on the estimated values of expected infections, expected reduction (with respect to the no-control case) and the expected number of removals using $\mathcal{H}$, $\mathcal{R}$ and $\mathcal{T}$. In Table 2 (ESM), we present the values of these estimates with their standard errors. Again, the performance of these measures is estimated on the same $m = 1000$ realisations of $(\theta, \underline{Q}) \sim \pi_0(\theta, \underline{Q}|\mathbf{y})$ ('pre-epidemics'). The minimum value of N' is taken to be 169 and 111 for Case (I) and (II) respectively, these values corresponding to the number of symptomatic individuals at t_{obs}.

Results indicate a greater difference in performance between the risk and threat measure than was observed for the uniformly distributed population. It can be seen from Figures 8 and 9 that, in general, prioritisation based on the threat map $\mathcal{T}$ is the most cost-effective control strategy in reducing the impact of the epidemics. This is particularly the case when resources are scarce (lower values of N') with the difference between results for the threat and risk measure decreasing as N' increases. The change in the discrepancy between threat and risk maps with increasing N' is most pronounced in Case (II), where the epidemic proceeds due to secondary infection only; for small values of N' the risk map's performance improves little on that of the hazard map but converges to that of the threat map as N' approaches its maximal value.

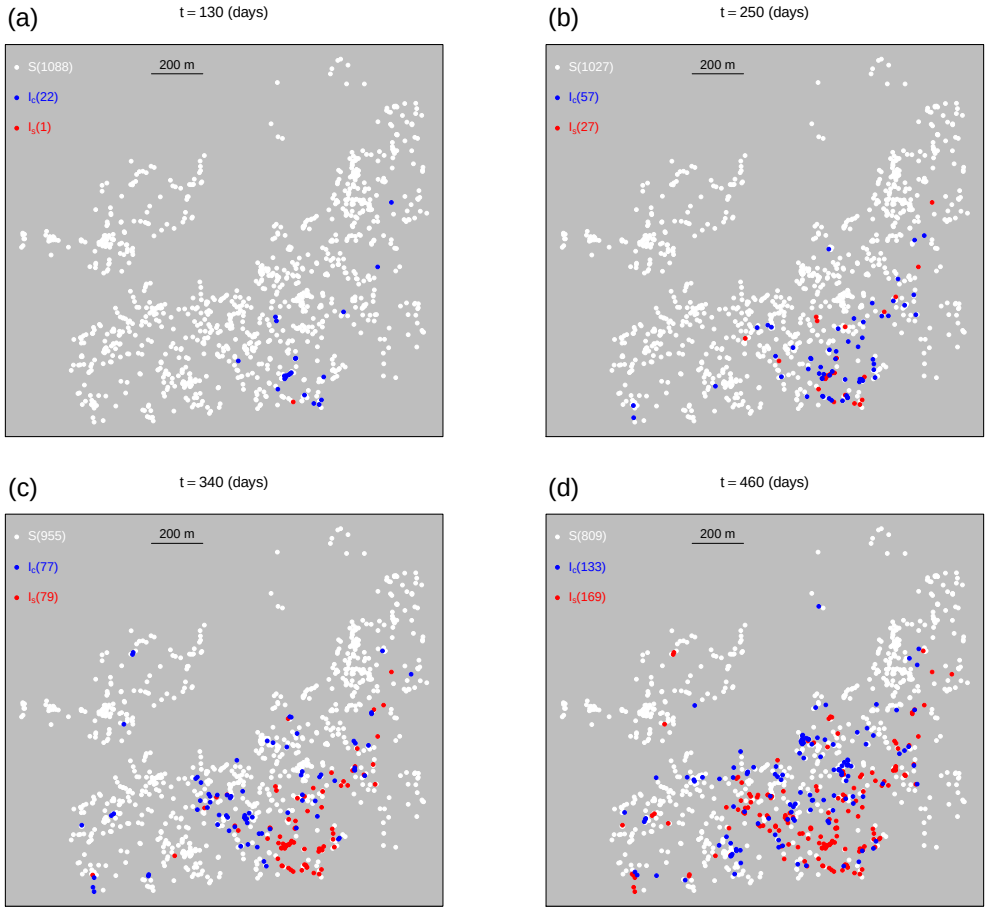

Figure 4: Case (I): with primary infection. A subset of a realisation of the disease progress maps made at 30-day intervals from $t = 130$ up to $t = 460$, on the citrus population of size $N = 1,111$ from a site located in Broward county. Only maps for $t = 130, 250, 340, 460$ are shown. Symptomatic hosts (I_s), cryptic infections (I_c) and susceptible hosts (S) at the time of the snapshot are denoted by red, blue and white dots respectively.

These results may be anticipated when one compares the threat and risk maps from Figures 6 and 7. For both Case (I) and Case (II) the hosts displaying the highest risk measures are located within the interior of the epidemic 'cluster' while those with the highest threat measure are located towards the periphery. It is to be expected that when N' is small the respective subsets selected using the risk and threat measures will be quite different and corresponding differences can be anticipated in the effectiveness of control.

The comparative performance of the threat and the risk measures, even for the clustered population, nevertheless depends on the range of the spatial kernel function. In Section 4 of the ESM we repeat the analysis of Case (1) presented in Figure 6, with kernel parameter $\alpha = 0.015, 0.04, 0.16, 0.2$ respectively, noting the smaller values of α imply a shorter range kernel. For this set of simulations we again see that the threat measure is markedly superior to the risk for smaller values of N' for $\alpha = 0.015, 0.04$ - particularly in the former case. However, when transmission is possible over longer ranges ($\alpha = 0.16, 0.2$), little difference in the performance of risk and threat is seen. This may be expected since, when transmission can occur over longer distances, the threat posed by an infection may be less sensitive to small-scale clustering in the epidemic and the susceptible population.

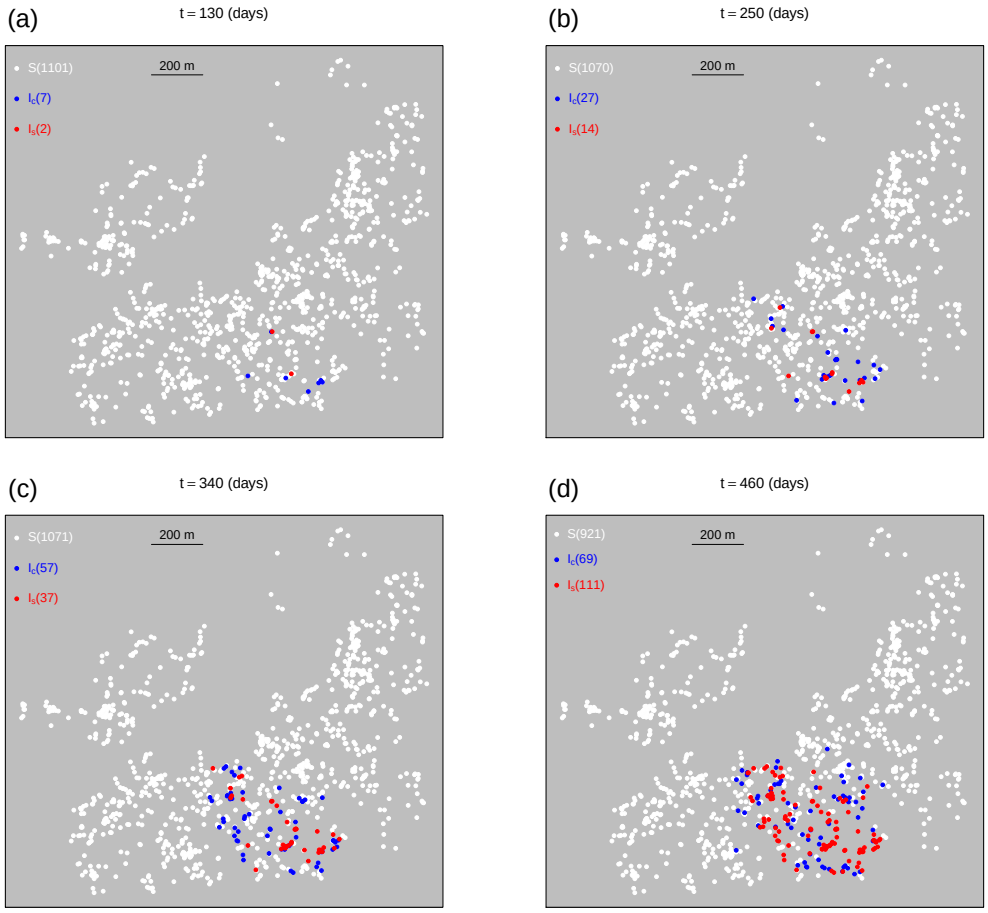

Figure 5: Case (II): without primary infection. A sample of a realisation of the disease progress maps made at 30-day intervals from $t = 130$ up to $t = 460$, on the citrus population of size $N = 1111$ from a site located in Broward county. Only maps for $t = 130, 250, 340, 460$ are shown. Symptomatic hosts (I_s), cryptic infections (I_c) and susceptible hosts (S) at the time of the snapshot are denoted by red, blue and white dots respectively. In comparison with Case (I), a far more clustered epidemic is observed.

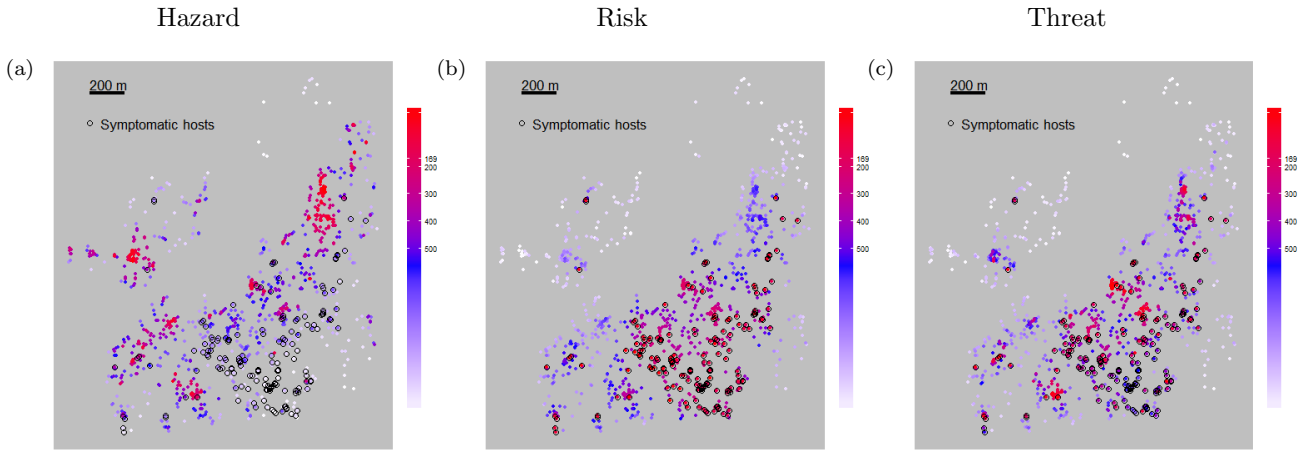

Figure 6: Posterior predictive maps of the hazard (a), risk (b) and threat (c) measures at $t_M = 460$ for Case (I) using equations (11- 13). The colour of points exhibits a gradation from white to blue to red with increasing values of the respective measure. The 169 symptomatic hosts detected during the survey are indicated by the black circles.

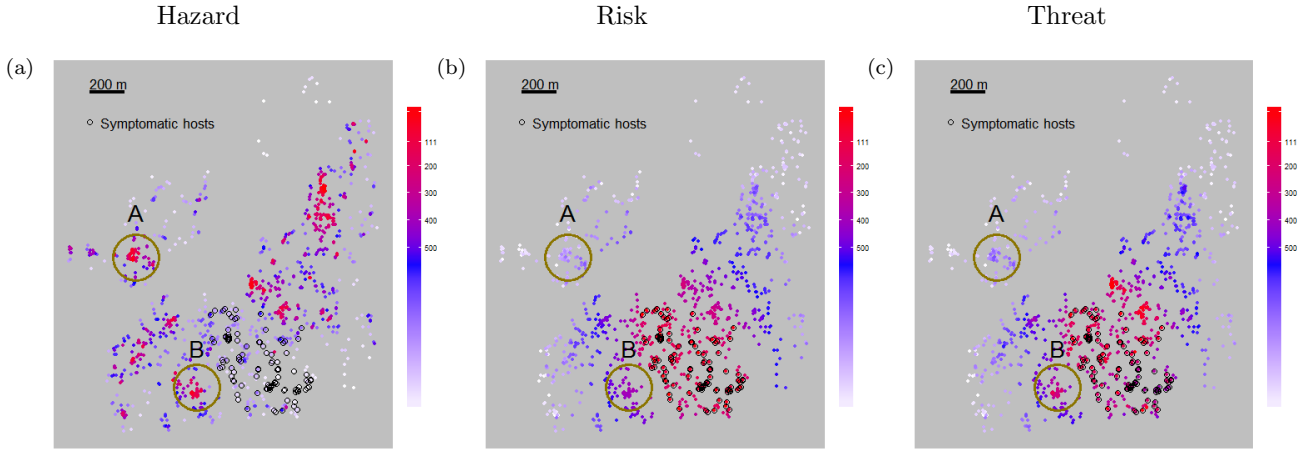

Figure 7: Posterior predictive maps of the hazard (a), risk (b) and threat (c) measures at $t_M = 460$ for Case (II) using equations (11-13). The colour of points exhibits a gradation from from white to red with increasing values of the respective measure. The 111 symptomatic hosts detected during the survey are indicated by the black circles. A cluster with intermediate risk (B) leads to high threat due to the high hazard while one with very low risk (A) ends up with relatively low threat even though the hazard is high.

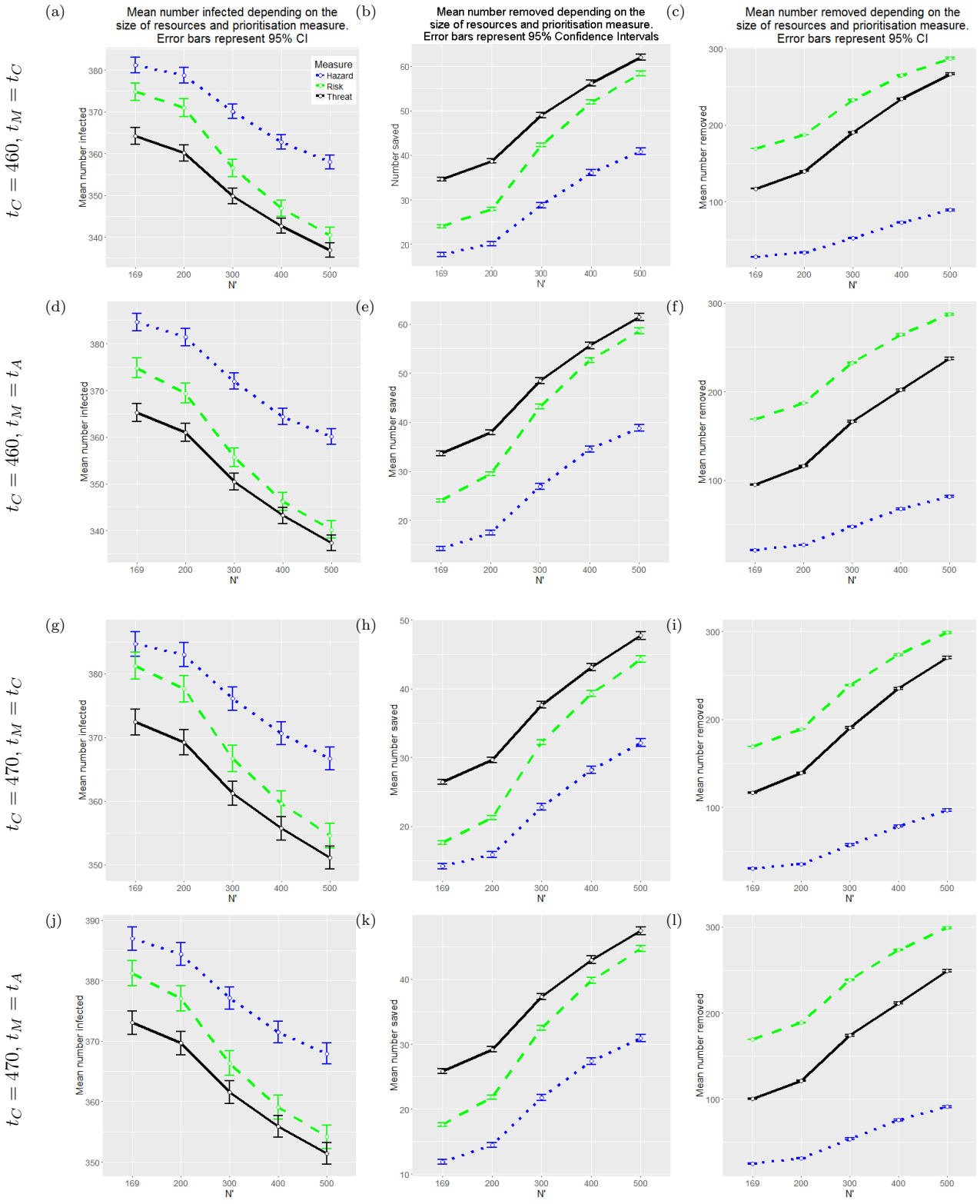

Figure 8: Marginal confidence intervals for the expected number of infections ((a), (d), (g), (j)), the estimated expected reduction in infections with respect to the no-control case ((b), (e), (h), (k)), and the expected number of removed hosts ((c),(f), (i), (l)) by $t_A = 500$ for Case (I) (primary infection). Results are presented for $t_C = 460$ and $t_C = 470$ using risk measures calculated from maps predicted at $t_M = t_C$ and $t_M = t_A$.

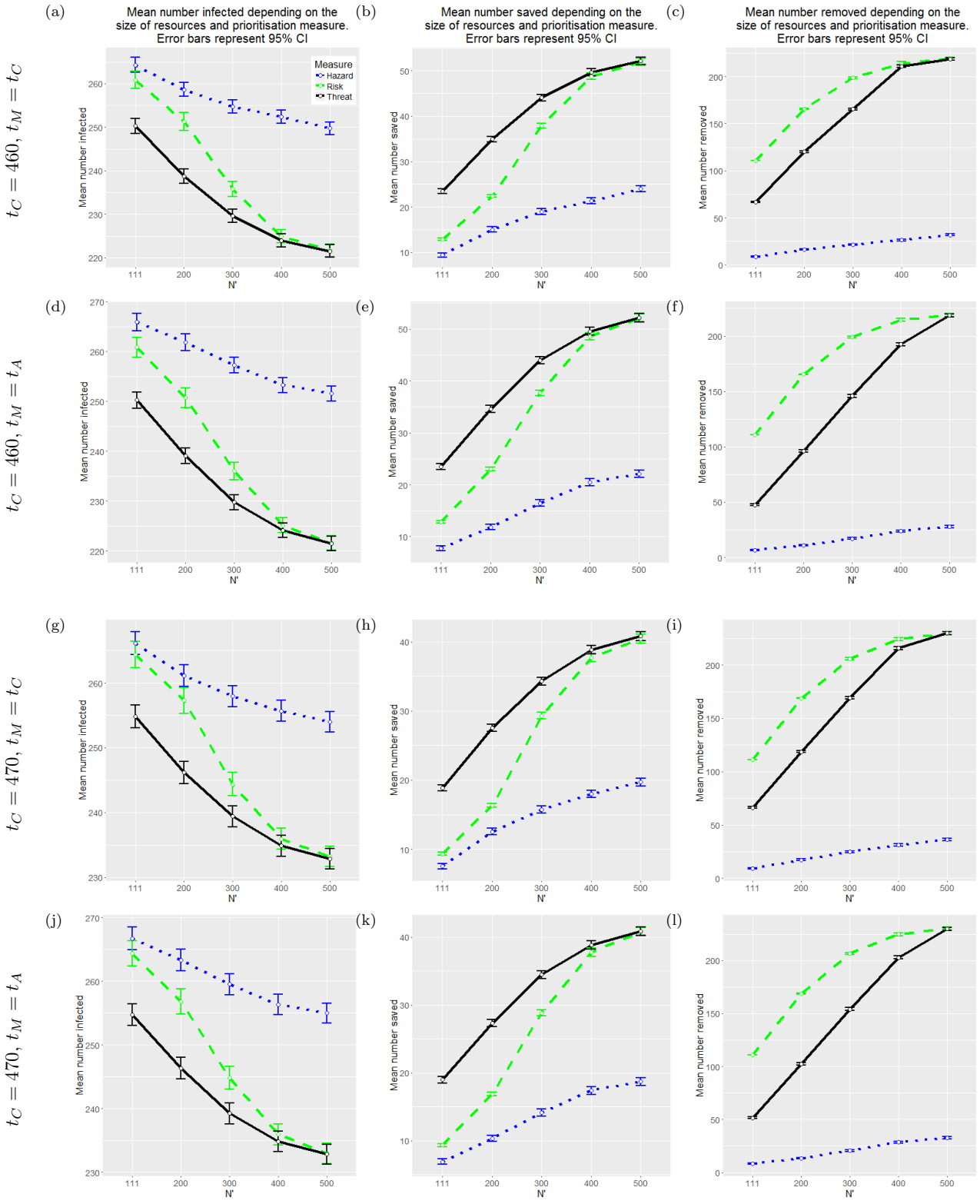

Figure 9: Marginal confidence intervals for the expected number of infections ((a), (d), (g), (j)), the estimated expected reduction in infections with respect to the no-control case ((b), (e), (h), (k)), and the expected number of removed hosts ((c),(f), (i), (l)) by $t_A = 500$ for Case (II) (primary infection). Results are presented for $t_C = 460$ and $t_C = 470$ using risk measures calculated from maps predicted at $t_M = t_C$ and $t_M = t_A$.

4 Discussion

The removal of infected hosts during the course of an epidemic is considered as the most efficient strategy for controlling epidemics of highly infectious diseases (Cook et al., 2008; Cunniffe et al., 2014). Therefore, when resources are scarce and the number of hosts that can be considered for removal is constrained, it is important that those hosts that may play the greatest role in the subsequent dynamics of the epidemic are targeted. This paper presents an efficient statistical computational framework to guide the targeting of control measures for highly infectious diseases with spatial dynamical transmission. In addition to formulating algorithms for model-based prediction of the efficacy of control strategies, we introduce a prioritisation scheme based on the idea that hosts with the highest threat - defined as the posterior expectation of the infectious challenge presented by a given host to susceptibles in the population - should be considered for removal first. For epidemics governed by SI dynamics, we use the computational methods to compare the threat-based prioritisation scheme with previously considered schemes.

An important feature of the computational approach is that it is embedded entirely in the Bayesian framework. This means that it is well suited to handling the challenges that often arise in epidemic modelling due to the partial nature of observations and allows unobserved quantities (here the precise times of infections) to be accommodated in analyses using data-augmentation. A second important feature is the use of functional-model representations of epidemics whereby the epidemic trajectory is represented as a deterministic function of the parameter vector and some latent stochastic process. This construction enables us to couple epidemics generated under various control strategies (Cook et al., 2008), by virtue of being driven by the same realisation of the latent process. In this paper we derive our latent process using the Sellke construction, which is easily handled within the MCMC and data augmentation methods that we use. Our results demonstrate that using the Sellke thresholds in this way induces strong positive correlation between the epidemic outcomes under alternative controls - leading to a reduction in the variance in the difference between outcomes under the controls.

The results presented here for the SI epidemic appear to suggest that the threat measure typically performs best out of the three measures considered. On the basis of the cases we have considered in our simulation study it appears that the superior performance of the threat measure is most pronounced when resources are scarce, in that only a small number of hosts can be considered for control, and when the epidemic spreads *via* short-range local transmission in a clustered host population. Under these conditions, the threat measure places high priority on hosts that are both likely to be infected and likely to have susceptible neighbours. Such hosts may be more likely to be located close to the edge of a clustered epidemic. Hosts that are likely to be infected but be largely surrounded by infected hosts are not prioritised so highly. The difference between the performance of the threat and the risk measures becomes less pronounced when the host population is uniformly distributed and when the range of the transmission kernel increases. Of course, in any practical scenario the likely performance of the measures considered (or alternative measures) should be investigated through studies akin to those carried out here, using observations of the emerging epidemic to be controlled. Nevertheless, the results support the notion that consideration of the threat measure for prioritising hosts is often a valuable strategy. Comparing the threat and risk measure in the context of Figure 9(a)-(c), we see that the expected reduction of epidemic size achieved using the threat map when $N' = 111$ would demand that $N' > 200$, were the same reduction to be achieved using the risk-based prioritisation scheme. At the same time the expected number of trees removed under the threat-based control for $N' = 111$ is less than half of that removed under the risk-based control achieving the same expected reduction. It should be noted that all the measures are posterior predictive expectations of unobserved functions of the epidemic trajectory and are, therefore, conditional on the observations available up to t_{obs}. It is not automatic that the same conclusions would emerge in the case where data were more or less extensive than is considered here and the quality of the posterior expectation as an estimator of the unobserved functions were improved or diminished as a consequence. Nevertheless, it makes intuitive sense that the threat measure should perform at least as well as the risk measure given that it targets those sites expected to present the greatest infectious challenge to susceptibles in the population.

For epidemics for which the SI model may not be appropriate we should not conclude that results obtained here, for example relating to the superiority of the threat measure, will automatically hold without further investigation. Nevertheless, the methods, and measures where appropriate, can be readily adapted to other settings in order to explore the relative merits of competing approaches to prioritising hosts for removal. Extensions of the basic SI model, such as the SEI, SIR, or SEIR models, can be accommodated within the computational framework. In the case of the SEIR model we may extend the latent-process vector $(\boldsymbol{\theta}, \underline{Q})$ to include vectors $\underline{T}_E, \underline{T}_I$, of sojourn times for each host in classes E and I. Given data y, we may use samples from $\pi(\boldsymbol{\theta}, \underline{Q}, \underline{T}_E, \underline{T}_I|\mathbf{y})$ (which can be readily obtained using MCMC methods) to couple the future trajectory of the epidemic under different control strategies involving host removal, as was done for the SI model using samples from $\pi(\boldsymbol{\theta}, \underline{Q}|\mathbf{y})$.

The range of prioritisation measures that can be defined will depend on the assumed model. For the SEIR

model, three versions of the risk measure considered here could be obtained by considering the posterior probability that a given host at time t_M is, respectively, in class E, in class I, or in $E \cup I$. For example, when the SI model is generalised to allow different infectivities β_c and β_s for cryptic and symptomatic hosts respectively, an appropriate threat measure could be defined as

$$\mathcal{T}_j(t_M) = E\left(\mathbb{1}_{\{x_j \le t_M - \Delta\}} \beta_s \sum_{i \neq j} K(d_{ij}, \alpha) \mathbb{1}_{\{x_i > t_M\}} + \mathbb{1}_{\{t_M - \Delta < x_j \le t_M\}} \beta_c \sum_{i \neq j} K(d_{ij}, \alpha) \mathbb{1}_{\{x_i > t_M\}} | \mathbf{y} \right) \tag{16}$$

and readily estimated using extensions of the MCMC methods. Equation (16) represents a measure that is composed of the sum of two separate components deriving from the cases where host j is in class I_S and I_C, respectively, at time t_M.

Ring-culling strategies (Tildesley et al., 2006; Cook et al., 2008; Neri et al., 2014; Cunniffe et al., 2015) can be assessed using the framework. In the SI model setting, for a given realisation $(\boldsymbol{\theta}, \underline{Q})$ it is straightforward to calculate the epidemic trajectory after t_{obs}, under the assumption that all hosts within distance r of a host, newly symptomatic at $t > t_{obs}$, are removed at time $t + \delta$, and to explore the impact of varying the culling-radius r and the response time δ.

The approach can be extended to alternative cost functions that incorporate economic factors, such as intervention costs (Forster and Gilligan, 2007; Neri et al., 2014) or cost of detection (Dybiec et al., 2004; Dybiec and Gilligan, 2005; Dybiec et al., 2009). For example it can accommodate the situation where diagnostic tests have imperfect sensitivity p and specificity q. This is achieved by augmenting the Sellke threshold for each host with a uniformly distributed random variable $z \sim U(0,1)$ (or a sequence of these when hosts may be tested multiple times) which determines the result of a diagnostic test, with sensitivity p and specificity q, applied to that host at a given time. If the host is susceptible at the time of the test then $z < q$ and $z \ge q$ result in negative and positive outcomes for the test. If the host is infected then $z < p$ and $z \ge p$ yield positive and negative outcomes respectively. This opens the way to explore, for example, the impact of using less sensitive, but less expensive, diagnostic tests on the efficacy of a control strategy.

We have considered the simple case whereby control strategies are selected on the basis of observations up to t_{obs}. Worthy of investigation is the potential gain in performance from allowing host prioritisations to be dynamic and adjustable in the light of new data obtained on the status of hosts already subjected to control.

It is not possible to pursue all the above challenges within the scope of this paper. Nevertheless, we are confident that the approach of using functional models and latent processes to couple epidemics under differing control regimes to estimate the efficacy of controls without excessive simulation is very appropriate for addressing them. A further, beneficial feature of the approach, which makes it robust to the increasing complexity arising from further developments of this nature, is the fact that any cost function is evaluated on a fixed set of parameter/latent process combinations meaning that computations are deterministic, once these combinations have been generated, and can be readily parallelised.

Author contributions

H.K.A., C.A.G., G.J.G designed research; H.K.A., G.S., N.J.C., T.R.G., C.A.G., G.J.G performed research; H.K.A., G.S., G.J.G performed mathematical and statistical analysis; H.K.A., G.S., N.J.C., T.R.G., C.A.G., G.J.G wrote the paper.

Acknowledgement

The authors are grateful to two anonymous referees for their helpful, constructive comments on an earlier version of this manuscript. Christopher Gilligan acknowledges the support of USDA, DEFRA and the Bill & Melinda Gates Foundation.

Data Accessibility

Data and C++ codes for testing method are uploaded at `http://people.ds.cam.ac.uk/ha411/Hola_Paper_interface.zip`.

Funding Statement

Hola Adrakey was supported during the course of this research by a James Watt Postgraduate Research Scholarship from Heriot-Watt University.

References

Boender GJ, Hagenaars TJ, Bouma A, Nodelijk G, Elbers AR, de Jong MC, van Boven, M (2007) Risk maps for the spread of highly pathogenic avian influenza in poultry. *PLoS Comput Biol*, 3:704–712.

Britton T, O'Neill P (2002) Bayesian inference for stochastic epidemics in populations with random social structure. *Scan J Stat*, 29:375–390.

Cook AR, Gibson GJ, Gottwald T, Gilligan CA (2008) Construction the effect of alternative intervention strategies on historic epidemics. *J R Soc Interface*, 5:1203–1213.

Cunniffe NJ, Laranjeira FF, Neri FM, DeSimone RE, Gilligan CA (2014) Cost-effective control of plant disease when epidemiological knowledge is incomplete: Modelling bahia bark scaling of citrus. *PLoS Comput Biol*, 10:e1003753.

Cunniffe NJ, Koskella B, Jessica E, Metcalf E, Parnell S, Gottwald TR, Gilligan CA (2015) Thirteen challenges in modelling plant diseases. *Epidemics*, 10:6–10.

Cunniffe NJ, Stutt RO, DeSimone RE, Gottwald TR, Gilligan CA (2015) Optimising and communicating options for the control of invasive plant disease when there is epidemiological uncertainty. *PLoS Comput Biol*, 11:e1004211.

Cunniffe NJ, Cobb RC, Meentemeyer RK, Rizzo DR, Gilligan CA (2016) Modeling when, where and how to manage a forest epidemic, motivated by sudden oak death in California. *PNAS*, 113:5640–5645.

DEFRA (2013) Chalara management plan. `http://www.defra.gov.uk/publications/`.

Dybiec B, Gilligan CA (2005) Opimising control of disease spread on networks. *Acta Phys Pol B*, 36:1509–1526.

Dybiec B, Kleczkowski A, Gilligan CA (2004) Controlling disease spread on networks with incomplete knowledge. *Phys Rev E*, 70:066145.

Dybiec B, Kleczkowski A, Gilligan CA (2009) Modelling control of epidemics spreading by long-range interactions. *J R Soc Interface*, 39:941–950.

Ferguson NM, Donnelly CA, Anderson RM (2001) The foot and mouth epidemic in Great Britain: Pattern of spread and impact of interventions. *Science*, 292:1155–1160.

Filipe JAN, Cobb RC, Meentemeyer RK, Lee CA, Valachovic YS, Cook AR, Rizzo DM, Gilligan CA (2012) Landscape Epidemiology and Control of Pathogens with Cryptic and Long-Distance Dispersal: Sudden Oak Death in Northern Californian Forests. *PLoS Comput Biol*, 8(1):e1002328.

Forster G, Gilligan CA (2007) Optimizing the control of disease infestations at the landscape scale. *PNAS*, 104:4984–4989.

Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem-US*, 81:2340–2361.

Gottwald TR, Graham JH, Schubert TS (2002) Citrus Canker: The Pathogen and its Impact. *Plant Heal Prog*, doi:10.1094/PHP-2002-0812-01-RV.

Gottwald TR, Sun X, Ripley T, Graham JH, Ferrandino F, Taylor EL (2002) Geo-Reference Spatiotemporal Analysis of the Urban Canker Epidemic in Florida. *Phytopathology*, 92:361–377.

Gottwald TR, Hughes G, Graham JH, Ripley T (2001) The citrus canker epidemic in Florida: The scientific Basis of regulatory eradication policy for an invasive species. *Phytopathology*, 91:30–34.

Graham JH, Gottwald TR, Cubero J, Achor DS (2004) Xanthomonas axonodis pv. citri: factors affecting successful eradiation of citrus canker. *Mol Plant Pathol*, 5.

Hyatt-Twynam SR, Stutt ROJH, Parnell S, Gottwald TR, Gilligan CA, Cunniffe NJ (2017) Risk-based management of invading plant disease. *New Phytol*.

Jewell CP, Kypraios T, Neal P, Roberts GO (2009) Bayesian analysis for emerging infectious diseases. *Bayesian Analysis*, 4:465–496.

Kao R (2003) The impact of local heterogeneity on alternative control strategies for foot and mouth disease. *Proc Biol Sci*, 270:2557–2564.

Lau MSY, Marion G, Streftaris G, Gibson GJ (2015) A systematic bayesian integration of epidemiological and genetic data. *PLoS Comput Biol*, 11:e1004633.

Neal P, Roberts G (2005) A case study in non-centering for data augmentation: stochastic epidemics. *Stat Comput*, 15:315–327.

Neri FM, Cook AR, Gibson GJ, Gottwald T, Gilligan CA (2014) Bayesian analysis for inference of an emerging epidemic: Citrus canker in urban landscapes. *PLoS Comput Biol*, 10:e1003587.

Parnell S, Gottwald T, van den Bosch F, Gilligan CA (2009) Optimal strategies for the eradication of asiatic citrus canker in heterogeneous host landscapes. *Phytopathology*, 99:1370–1376.

Parry M, Gibson GJ, Parnell S, Gottwald TR, Irey MS, Gast TC, Gilligan CA (2014) Bayesian inference for an emerging arboreal epidemic in the presence of control. *PNAS*, 111:6258–6262.

Schubert TS, Rixvi SA, Sun X, Gottwald TR, Graham JH, Dixon WN (2001) Meeting the challenge of eradicating citrus canker in florida again. *Plant Dis*, 85:340–356.

Sellke T (1983) On the asymptotic distribution of the size of a stochastic epidemic. *J Appl Probab*, 20:390–394.

te Beest DE, Hagenaars TJ, Stegeman JA, Koopmans MP, van Boven M (2011) Risk based culling for highly infectious diseases of livestock. *Vet Res*, 42:81.

Thompson D, Murriel P, Russell D, Osborne P, Bromley M, Creigh-Tyte S, Brown C (2004) Economic costs of the foot and mouth disease outbreak in the United Kingdom in 2002. *Rev Sci Tech*, 21:675–687.

Tildesley MJ, Savill NJ, Shaw DJ, Deardon R, Brooks SP, Woolhouse MEJ, Grenfell BT, Keeling MJ (2006) Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. *Nature*, 440:83–86.

Tildesley MJ, Bessell PR, Keeling MJ, Woolhouse ME (2009) The role of pre-emptive culling in the control of foot and mouth disease. *Proc R Soc B*, 276:3239–3248.

Thompson RN, Cobb RC, Gilligan CA, Cunniffe NJ (2016) Management of invading pathogens should be informed by epidemiology rather than administrative boundaries. *Ecological Modelling*, 324:28–32.

USDA/APHIS in consultation with the Florida citrus industry and other stakeholders (2006) Citrus health response program (chrp) minimum standards for citrus health in florida. `http://www.aphis.usda.gov/plant_health/plant_pest_info/citrus/downloads/chrp.pdf`. Version 1.0.