Appl. Statist. (2018)
67, *Part* 4, *pp.* 1–28

Optimal treatment allocations in space and time for on-line control of an emerging infectious disease

Eric B. Laber, Nick J. Meyer, Brian J. Reich and Krishna Pacifici

North Carolina State University, Raleigh, USA

Jaime A. Collazo

US Geological Survey North Carolina Cooperative Fish and Wildlife Research Unit, and North Carolina State University, Raleigh, USA

and John Drake

University of Georgia, Athens, USA

[*Read before* The Royal Statistical Society *on Wednesday, March 14th, 2018,* Professor R. Henderson *in the Chair*]

Summary. A key component in controlling the spread of an epidemic is deciding where, when and to whom to apply an intervention. We develop a framework for using data to inform these decisions in realtime. We formalize a treatment allocation strategy as a sequence of functions, one per treatment period, that map up-to-date information on the spread of an infectious disease to a subset of locations where treatment should be allocated. An optimal allocation strategy optimizes some cumulative outcome, e.g. the number of uninfected locations, the geographic footprint of the disease or the cost of the epidemic. Estimation of an optimal allocation strategy for an emerging infectious disease is challenging because spatial proximity induces interference between locations and the number of possible allocations is exponential in the number of locations, and because disease dynamics and intervention effectiveness are unknown at outbreak. We derive a Bayesian on-line estimator of the optimal allocation strategy that combines simulation–optimization with Thompson sampling. The estimator proposed performs favourably in simulation experiments. This work is motivated by and illustrated using data on the spread of white nose syndrome, which is a highly fatal infectious disease devastating bat populations in North America.

Keywords:

1. Introduction

Dynamical systems on networks and more general spatial domains have proved to be an effective modelling tool in many areas of science (Strogatz, 2001). Applications include ecological food webs (Williams and Martinez, 2000), electrical power grids, cellular and metabolic networks (Kohn, 1999), the World Wide Web (Broder *et al.*, 2000) and human mobility (Truscott and Ferguson, 2012). Interest in network anatomy and function underlies the greater movement towards research on complex systems. Recently, there has been increased interest in trying to understand dynamical systems on networks with the objective of administering control over

Address for correspondence: Eric B. Laber, Department of Statistics, North Carolina State University, 5216 SAS Hall, 2311 Stinson Drive, Raleigh, NC 27695-8203, USA.
E-mail: laber@statincsu.edu

 0035–9254/18/67000

some process evolving over the network. An important example is the control of an epidemic evolving on a network of individuals. Fatal emerging diseases such as West Nile virus (Kilpatrick, 2011), white nose syndrome (WNS) (Maher *et al*., 2012), foot-and-mouth disease (Tildesley *et al*., 2006), and severe acute respiratory syndrome (Hufnagel *et al*., 2004) represent a serious threat to ecological and environmental systems and to human health. The effect of bat loss due to WNS is projected to produce several billions of dollars of agricultural costs per year (Subcommittee on Fisheries, Wildlife, and Oceans, 2011). Understanding the dynamics of these epidemics and providing tools to control them efficiently and effectively are of paramount importance.

A key component in controlling the spread of an epidemic is deciding where, when and to whom to apply an intervention. A treatment allocation strategy formalizes this process as a sequence of functions, one per treatment period, that map up-to-date information on the epidemic to a subset of locations to receive treatment. An optimal treatment allocation strategy optimizes the expectation of some cumulative outcome, e.g. the cumulative number of infected individuals, the geographic footprint of the disease, the estimated total cost of the disease or a composite of several important outcomes. Estimation of an optimal treatment allocation strategy for an emerging epidemic presents several major challenges:

(a) scarcity of data—at the onset of the epidemic there is little information about disease dynamics and typically no information on the effectiveness of potential treatments;
(b) scalability—the number of possible allocations is exponential in the number of locations; for example, in the problem of WNS, there are more than 1100 locations leading to more treatment allocations than can possibly be enumerated by using existing computing resources;
(c) interference—dependence between locations violates the no interference between experimental units assumption (Sobel, 2006; Hudgens and Halloran, 2008);
(d) a long time horizon—an epidemic can persist for decades before eradication, and thus an optimal treatment allocation strategy must adapt to evolving logistical constraints, technologies and system dynamics.

We propose an on-line estimator of the optimal treatment allocation strategy that is designed to overcome these challenges. At each time point, the method proposed draws a model from the posterior distribution over system dynamics models and the estimated optimal allocation strategy is the maximizer over a prespecified class of strategies of the mean outcome under this model. The system dynamics model and estimated optimal allocation strategy are updated each time that new data are collected to provide a continually evolving strategy. Furthermore, the class of potential allocation strategies is chosen to reduce computational complexity when scaling to large decision problems and to ensure that logistical or feasibility constraints are satisfied. We show that the estimator proposed can scale to problems with more than 1000 nodes, four covariates per node, 15 treatment periods and $O(10^{150})$ possible allocations at each time period.

The methodology proposed is related to the idea of a dynamic treatment regime in personalized medicine. A dynamic treatment regime is a sequence of decision rules, one per treatment stage, that map up-to-date patient information to a recommended treatment (Murphy, 2003a; Robins, 2004a; Schulte *et al*., 2014). Thus, like a treatment allocation strategy, a dynamic treatment regime is a sequence of functions that is used to dictate treatment decisions over time. Furthermore, one approach to estimating a dynamic treatment regime is to model the mean outcome as a function of each regime in a prespecified class and then to take the maximizer over this class as the estimated optimal regime (Robins *et al*., 2008; Orellana *et al*., 2010; Zhao *et al*., 2012; Zhang *et al*., 2012a,b, 2013; Zhao *et al*., 2014a,b; Kang *et al*., 2014). However, despite

these similarities, the challenges that were mentioned previously prevent direct application of methodology for dynamic treatment regimes to the problem of spatiotemporal treatment allocation. Methods for dynamic treatment regimes assume that the data comprise a large number of independent and identically distributed trajectories observed over time. In contrast, in the allocation problem, we observe a single observation over the spatial domain at each time point; hence there is no independent replication. Furthermore, existing methods for dynamic treatment regimes are designed for settings with a small number of treatment options at each treatment stage, e.g. between two and five, whereas, in the spatial allocation problem, there are an astronomically large number of potential treatments. There has been some research on continuous treatments in dynamic treatment regimes (Rich, 2013; Rich *et al*., 2014; Laber and Zhao, 2015); however, these methods heavily rely on smoothness of an outcome regression model across treatment values which does not apply in the treatment allocation problem.

Both estimation of dynamic treatment regimes and estimation of an optimal treatment allocation fall under the umbrella of reinforcement learning problems (Bertsekas, 1996; Sutton and Barto, 1998; Powell, 2007; Sugiyama, 2015). Our proposed estimator is an approximate variant of Thompson sampling (Thompson, 1933) wherein allocations are chosen with probability that is proportional to the posterior probability that they are optimal. Thompson sampling has been studied in the reinforcement learning literature primarily in its application to bandit problems (Scott, 2010; Chapelle and Li, 2011; Agrawal and Goyal, 2011, 2012, 2013; Kaufmann *et al*., 2012; Korda *et al*., 2013; Gopalan *et al*., 2014; Russo and Van Roy, 2014). Osband *et al*. (2013) and Gopalan and Mannor (2015) applied Thompson sampling to sequential decision problems modelled as Markov decision processes. However, these estimators require

(a) a finite set of system states, and
(b) that a fixed allocation strategy be applied without adjustment for potentially long periods of time.

In the settings that we consider, the system state is continuous and high dimensional (making discretization impractical) and the application of a fixed suboptimal allocation strategy for a prolonged period is neither ethical nor feasible. For a comprehensive survey of Bayesian reinforcement learning see Ghavamzadeh *et al*. (2015).

The work proposed adds to the large literature on mathematical spatial–transmission models for disease modelling and control. Within this literature, a common approach is to postulate a mathematical model of disease spread and then to use simulation experiments to evaluate and compare candidate intervention strategies (see Anderson *et al*., (1992), Hufnagel *et al*. (2004), Riley (2007), Hollingsworth (2009), Ma *et al*. (2009), Keeling and Rohani (2011) and references therein). These models have generated new insights into disease transmission and control strategies across a wide range of application domains including avian influenza (Le Menach *et al*., 2006; Jung *et al*., 2009), Chagas disease (Barbu *et al*., 2009, 2011), Ebola virus disease (Lekone and Finkenstädt, 2006; Kramer *et al*., 2016; Li *et al*., 2017); foot-and-mouth disease (Ferguson *et al*., 2001a, b; Keeling, 2005; Tildesley *et al*., 2006), human immunodeficiency virus and acquired immune deficiency syndrome (Jacquez *et al*., 1988; Korenromp *et al*., 2000), severe acute respiratory syndrome (Huang *et al*., 2004; Bauch *et al*., 2005; Hollingsworth *et al*., 2006) and smallpox (Kaplan *et al*., 2002; Bozzette *et al*., 2003; Ferguson *et al*., 2003; Kretzschmar *et al*., 2004), among others. The proposed Thompson sampling estimator relies on a working model for the underlying disease dynamics and thereby benefits from this rich literature on disease modelling. However, the estimator proposed is distinct from these approaches in that it

(a) considers on-line estimation of an optimal allocation strategy which requires balancing

the choice of treatment allocations that lead to maximal model improvement with the choice of treatment allocations that are optimal under the current estimated model,
(b) optimizes over a large (possibly infinite) class of allocation strategies and
(c) accommodates evolving resource and logistical constraints.

Thus, where much of the focus of mathematical disease modelling has been on building and validating high quality transmission models, our focus is on how to incorporate these models in optimal on-line treatment allocation.

In Section 2, we discuss one of the motivating problems for the work proposed: controlling the spread of WNS in bats. In Section 3, we formally define an optimal treatment allocation strategy by using potential outcomes and discuss the problem of interference. In Section 4, we develop our estimator of the optimal treatment allocation strategy and construct a class of strategies that are flexible but highly scalable. In Section 5, we evaluate the performance of the method by using simulation experiments. In Section 6, we apply the methodology to data on the spread of WNS. Future work and open problems are discussed in Section 7.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

`http://www.rss.org.uk/preprints`

2. White nose syndrome in bats

WNS is a disease that is caused by the fungus *Pseudogymnoascus destructans* (formerly *Geomyces destructans*) and predominately affects hibernating bats in North America (Blehert *et al*., 2009). An infected bat will present with a white fungus on its muzzle, ears and/or wings, and erratic behaviour during hibernation. The erratic behaviour during hibernation depletes fat reserves and expends valuable energy, resulting in low survival and death (Blehert *et al*., 2009). Mortality rates exceed 90% in some areas and more than 5.7 million bats have died because of WNS (Blehert *et al*., 2009; US Fish and Wildlife Service, 2015).

WNS was first recorded in Schoharie County, New York State, in 2006 (Blehert *et al*., 2009) and is now found in 25 states, five Canadian provinces, as far south as Mississippi, and as far west as Missouri; Fig. 1. More than half of the 47 species of bats in the USA hibernate, making them vulnerable to exposure. Currently, two endangered species, the Gray bat, *Myotis grisescens*, and the Indiana bat, *Myotis sodalis*, as well as one threatened species, the Northern Long-eared bat, *Myotis septentrionalis*, are infected with WNS (Blehert *et al*., 2009). The ecological damage due to loss of bats and the speed of spread are unprecedented and the long-term damage is still considered to be immeasurable (Blehert *et al*., 2009). Short-term estimates of economic damage hover around \$3.7 billion year^{-1} mainly due to agricultural loss (Boyles *et al*., 2011). The estimated value of bats to the entire agricultural industry is \$22.9 billion year^{-1} not including many secondary effects and impacts, e.g. downstream effects of increased use of pesticides; predation effects on evolved resistance of insects to pesticides and genetically modified crops (Boyles *et al*., 2011).

Because of the economic and ecological effects, there is a tremendous need for a comprehensive national plan to control the spread of WNS. In 2009, the US Fish and Wildlife Service, along with state agencies and universities, convened to create a national plan for the control of WNS (US Fish and Wildlife Service, 2015). This plan outlines necessary actions for co-ordination and provides an overall template to prevent further spread of WNS. Although the national plan puts forth the first steps in co-ordination between states and other agencies, it does not explicitly provide a treatment plan or strategy to control WNS (Szymanski *et al*., 2009). Each state is left to

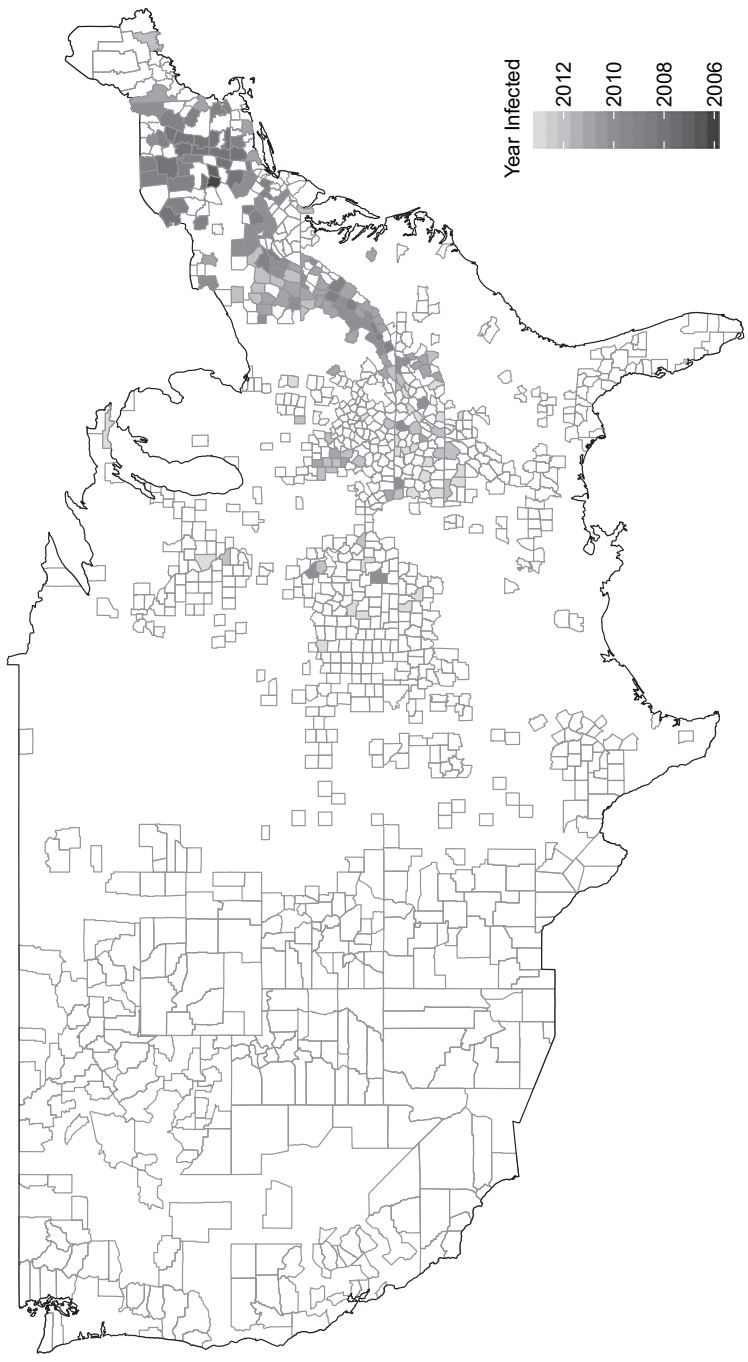

Fig. 1. Spread of WNS (US Fish and Wildlife Service, 2015): outlined counties contain caves; those without colour were uninfected at June 2014

implement treatments at its own discretion. Potential treatments include antifungal biological or non-chemical agents for bats at risk, modifying cave environmental variables, e.g. temperature and humidity, to slow fungus growth and to improve bat survival, vaccines to boost resistance and artificial caves (Cornelison *et al*., 2014; Hoyt *et al*., 2015). Unfortunately, many of these have not been tested in the field and their efficacy is currently unknown. Additional challenges exist because the disease has a highly complex nature of spread including a large spatial range (Maher *et al*., 2012). Therefore, to maximize what benefits these treatments may provide, it is essential to develop a principled, adaptive and data-driven control strategy that addresses the full potential range of WNS before further devastation occurs. We estimate such a control strategy and demonstrate that, if implemented, it may have a profound effect on the course of the current epidemic.

3. Defining an optimal treatment allocation strategy

We consider a decision problem evolving over a countably infinite set of treatment periods and a finite number of locations. The locations may represent physical locations in space, e.g. parcels of land identified as candidates for an intervention, or the locations may be nodes in a network, e.g. individuals in a social network. In the application to WNS, the data are provided at the county level, and thus cave bearing counties compose the locations of interest. At each time point, a decision maker observes information describing the current state of each location and subsequently uses this information to decide which locations should receive treatment. In the control of epidemics, location information would include information on the spread of the disease, e.g. infection status and time since infection among the infected, as well as features that are related to susceptibility or contagiousness. In WNS, for each county we observe the infection status, time since infected, number of caves, average winter temperature and a measure of species richness. For simplicity, our development considers the setting in which there are two possible choices at each location: apply a treatment or do nothing. However, the methodology proposed can be extended to handle settings in which several treatment options are available at each location. A treatment allocation strategy formalizes the treatment allocation process as a map from current information on all locations to a probability distribution over possible allocations. An allocation strategy is said to be optimal if it maximizes the mean cumulative utility over a prespecified class of strategies (minimizing cost can be handled in the obvious way).

Let $\mathcal{L} = \{1, \ldots, L\}$ denote the set of locations and $\mathcal{T} = \{1, 2, \ldots\}$ the set of treatment stages. The treatment stages may be dictated by the evolving decision process. Define $\mathbf{S}_l^t \in \mathbb{R}^p$ to be a summary of the information that is collected at location $l \in \mathcal{L}$ up to and including time $t \in \mathcal{T}$ and let $\mathbf{S}^t$ be $\{\mathbf{S}_l^t\}_{l \in \mathcal{L}}$; we assume that $\mathbf{S}^t$ is completely observed and measured without error. Let $A_l^t \in \{0, 1\}$ denote an indicator that location l received treatment at time t and $\mathbf{A}^t = \{A_l^t\}_{l \in \mathcal{L}}$ is the allocation at time t. Let $\mathcal{B}_L$ denote the set of all probability distributions over $\{0, 1\}^L$; and for a random variable U write $\text{supp}(U)$ to denote the support of U. A treatment allocation strategy π is a function from $\mathcal{S} = \text{supp}(\mathbf{S}^t)$ into $\mathcal{B}_L$ so that, under π, a decision maker who is presented with $\mathbf{S}^t = \mathbf{s}^t$ will select allocation $\mathbf{a}^t$ with probability $\pi(\mathbf{a}^t; \mathbf{s}^t)$. Allocation strategies of this type are termed stochastic strategies to contrast them with deterministic strategies which map states to allocations rather than to a distribution over allocations (Sutton and Barto, 1998). In the context of on-line estimation and optimization, the use of stochastic allocation strategies is critical to ensure consistent estimation of an optimal strategy (Kaelbling *et al*., 1996; Cesa-Bianchi and Lugosi, 2006) much in the same way that randomization is critical in adaptive clinical trials to ensure consistent estimation of an optimal treatment (Berry and Fristedt, 1985); see the on-line supplemental materials for an illustrative example. Let $Y_l^t \in \mathbb{R}$ denote an outcome

measured at location l at time t and let $\mathbf{Y}^t = \{Y_l^t\}_{l\in\mathcal{L}}$. For a prespecified constant $\gamma \in (0,1)$, the goal is to choose an allocation strategy that maximizes the mean of the discounted total utility $\Sigma_{t\geqslant 1}\gamma^{t-1}u(\mathbf{Y}^t)$, where $u(\cdot)$ is a scalar utility function and the constant γ balances proximal and distal outcomes. In some settings, it may be desirable to choose an alternative measure of cumulative utility, e.g. $\lim_{T\to\infty} T^{-1}\Sigma_{t=1}^{T} u(\mathbf{Y}^t)$; our methodology can be directly extended to handle such alternatives. We formalize the notion of an optimality allocation strategy by using potential outcomes (Rubin, 1978; Splawa-Neyman *et al.*, 1990).

Let Π denote a class of allocation strategies of interest; throughout, we implicitly assume that all allocation strategies under consideration belong to Π. Hence, the definition of optimality depends on Π. This class can be used to enforce logistical constraints, e.g. a limit on the number of locations that can be treated at each time point. Because our estimation algorithm is online, this class of allocation strategies can be changed in realtime to reflect changing constraints. Define $\mathcal{F} = \{\mathbf{a} \in \{0,1\}^L : \mathbf{a} \in \text{supp}(\pi) \text{ for some } \pi \in \Pi\}$ to be the set of feasible allocations. We use overline notation to denote past history, e.g. $\bar{\mathbf{a}}^t = \{\mathbf{a}^v\}_{v=1}^t$, and an asterisk superscript to denote potential outcomes. For example, $Y^{*t}(\bar{a}^t)$ denotes the outcome that would be observed under treatment sequence $\bar{a}^t$. Define $\mathbf{W}^* = \{\mathbf{Y}^{*t}(\bar{\mathbf{a}}^t), \mathbf{S}^{*t+1}(\bar{\mathbf{a}}^t) : \mathbf{a}^t \in \mathcal{F}\}_{t\in\mathcal{T}}$ to be the set of potential outcomes under $\{\mathbf{a}^t\}_{t\in\mathcal{T}}$, i.e. the states and outcomes that would be observed under actions $\{\mathbf{a}^t\}_{t\in\mathcal{T}}$, where we have defined $\mathbf{S}^{*1}(\bar{\mathbf{a}}^0) \equiv \mathbf{S}^1$ for convenience.

For any $\pi \in \Pi$, let $\{\xi_\pi^t(\mathbf{s})\}_{t\in\mathcal{T},\mathbf{s}\in\mathcal{S}}$ denote a collection of independent random variables so that $P\{\xi_\pi^t(\mathbf{s}^t) = \mathbf{a}^t\} = \pi(\mathbf{a}^t; \mathbf{s}^t)$. Define

$$\mathbf{Y}^{*t}(\pi) \triangleq \sum_{\bar{\mathbf{a}}^t} \mathbf{Y}^{*t}(\bar{\mathbf{a}}^t) \prod_{v=1}^{t} \mathcal{I}[\xi_\pi^v\{\mathbf{S}^{*v}(\bar{\mathbf{a}}^{v-1})\} = \bar{\mathbf{a}}^v]$$

to be the potential outcome under allocation strategy π, where $\mathbf{S}^{*1}(\bar{\mathbf{a}}^0) = \mathbf{S}_1$. An allocation strategy $\pi^{\text{opt}} \in \Pi$ is optimal if

$$\mathbb{E}\Big[\sum_{t\geqslant 1}\gamma^{t-1}u\{\mathbf{Y}^{*t}(\pi^{\text{opt}})\}\Big] \geqslant \mathbb{E}\Big[\sum_{t\geqslant 1}\gamma^{t-1}u\{\mathbf{Y}^{*t}(\pi)\}\Big]$$

for all $\pi \in \Pi$. If there are multiple optimal strategies within Π there is no loss by choosing between them arbitrarily. Thus, for brevity, we assume hereafter that π^{opt} is unique. To estimate π^{opt} from the observed data, we require assumptions about the data-generating mechanism. At time t, the available data to estimate π^{opt} are $\mathbf{H}^1 = \mathbf{S}^1$ if $t = 1$ and $\mathbf{H}^t = (\mathbf{S}^1, \mathbf{A}^1, \mathbf{Y}^1, \ldots, \mathbf{S}^{t-1}, \mathbf{A}^{t-1}, \mathbf{Y}^{t-1}, \mathbf{S}^t)$ if $t \geqslant 2$. We make the following assumptions.

Assumption 1 (sequential ignorability (Robins, 2004a)): $\mathbf{A}^t \perp\!\!\!\perp \mathbf{W}^* | \mathbf{H}^t$ for all $t \in \mathcal{T}$.

Assumption 2: the observed outcomes are the potential outcomes under treatment actually received, $\mathbf{Y}^t = \mathbf{Y}^{*t}(\bar{\mathbf{A}}^t)$ and $\mathbf{S}^t = \mathbf{S}^{*t}(\bar{\mathbf{A}}^{t-1})$ for all $t \in \mathcal{T}$.

Assumption 3 (positivity): there exists $\epsilon > 0$ so that $P(\mathbf{A}^t = \mathbf{a} | \mathbf{H}^t) > \epsilon$ for all $\mathbf{a} \in \mathcal{F}$ and $t \in \mathcal{T}$ with probability 1.

Although we have stated assumption 2 as an assumption, there is some debate about whether this should instead be taken as an axiom (Pearl, 2010; VanderWeele *et al.*, 2013; Keele, 2015); in addition, we implicitly assume throughout that there are no hidden forms of treatment. Given a data-generating process which satisfies assumptions 1–3, for any $\pi \in \Pi$ it follows that

$$\mathbb{E}\Big[\sum_{t\geqslant 1}\gamma^{t-1}u\{\mathbf{Y}^{*t}(\pi)\}\Big] = \lim_{T\to\infty}\int\Big\{\sum_{t=1}^{T}\gamma^{t-1}u(\mathbf{y}^t)\Big\}\prod_{v=1}^{T}\{f_v(\mathbf{y}^v|\mathbf{h}^v, \mathbf{a}^v)\pi(\mathbf{a}^v; \mathbf{s}^v) \times g_v(\mathbf{h}^v|\mathbf{h}^{v-1})\}\mathrm{d}\lambda(\bar{\mathbf{y}}^T, \bar{\mathbf{a}}^T, \bar{\mathbf{h}}^T), \quad (1)$$

where f_v is the conditional density for $\mathbf{Y}^v$ given $\mathbf{H}^v$ and $\mathbf{A}^v$, g_v is the conditional density for $\mathbf{H}^v$ given $\mathbf{H}^{v-1}$ with $g_1(\mathbf{h}^1|\mathbf{h}^0) = g_1(\mathbf{h}^1)$ and λ is a dominating measure (in our applications, this will be a product of Lebesgue and counting measures corresponding to the continuous and discrete components of $\bar{\mathbf{Y}}^T$, $\bar{A}^T$ and $\bar{H}^T$). Thus, result (1) shows how the expected cumulative utility can be expressed by using the data-generating model.

The foregoing assumptions along with the assumption of no interference between experimental units are standard in causal inference for non-spatial sequential decision-making problems (Chakraborty and Moodie, 2013; Schulte *et al.*, 2014). However, in spatiotemporal decision problems, the proximity of the locations can induce spillover effects thereby causing interference between experimental units (locations) (Halloran and Struchiner, 1995; Diez Roux, 2004; Hong and Raudenbush, 2006; Hudgens and Halloran, 2008; VanderWeele and Tchetgen Tchetgen, 2011; Ogburn and VanderWeele, 2014). Furthermore, in many settings, there are cost constraints of the form $\Sigma_{l\in\mathcal{L}} v_l^t a_l^t \leqslant c_t$, where v_l^t is the cost of applying treatment at location l at time t and c_t is a total budget at time t. Constraints of this form are another reason why the decision to treat one location requires consideration of all others, i.e. applying treatment at one location reduces the available budget for applying treatments elsewhere. Standard methods for estimating optimal decision rules, e.g. Q- and A-learning (Murphy, 2003b, 2005; Robins, 2004b; Blatt *et al.*, 2004; Goldberg and Kosorok, 2012; Schulte *et al.*, 2014; Laber *et al.*, 2014) and policy search (Robins *et al.*, 2008; Orellana *et al.*, 2010; Zhang *et al.*, 2012a,b, 2013; Zhao *et al.*, 2012, 2014a,b), are based on independent application of treatment to each unit. Thus, to apply these methods without additional assumptions, it would be necessary to treat the collection of all locations as a single experimental unit. In this case, there are $O(2^L)$ available allocations at each time and a single observation available for estimation. Furthermore, existing methods rely on estimation of part or all of the conditional distribution of $\mathbf{Y}^t$ given $(\bar{\mathbf{S}}^t, \bar{\mathbf{A}}^t)$ treating $\bar{\mathbf{A}}^t$ as a categorical variable with 2^L levels. Fitting such a model, even if sufficient replications were available to identify the distribution, would be computationally infeasible.

4. Estimating an optimal allocation strategy

In the context of an emerging epidemic, there is typically little or no data that can be used to form reliable estimators for some (or all) components of the system dynamics model. Thus, it is essential to add information from scientific theory to the estimation process. We integrate scientific theory to the estimation process by taking a Bayesian perspective on parameter uncertainty and allowing the use of informative priors on some (or all) of the parameters indexing our postulated system dynamics model.

An overview of our estimation procedure is as follows. Let $\mathcal{D}$ denote a class of deterministic, i.e. non-stochastic, allocation strategies. Under $d \in \mathcal{D}$, a decision maker who is presented with state $\mathbf{S} = \mathbf{s}$ will select allocation $d(\mathbf{s})$. At each time t, we draw a system dynamics model from the posterior distribution over dynamics models and subsequently use simulation–optimization (Law *et al.*, 1991; Banks *et al.*, 1998; Gosavi, 2003) to compute a maximizer, say $\hat{d}^t$, of equation (1) over $\mathcal{D}$ where equation (1) is computed with respect to the sampled dynamics model. Given state $\mathbf{S}^t = \mathbf{s}^t$, the selected allocation at time t is $\hat{d}^t(\mathbf{s}^t)$. This implicitly defines a stochastic allocation, $\hat{\pi}^t(\mathbf{s}^t)$, as a mixture over $\{d(\mathbf{s}^t) : d \in \mathcal{D}\}$ with mixture probabilities equal to the posterior probability that d is the maximizer of equation (1); thus, the implied class of stochastic strategies, Π, is the class of all mixtures over strategies in $\mathcal{D}$. A schematic diagram for this procedure is displayed in Fig. 2(a).

This approach can be viewed as a version of Thompson sampling wherein allocations are

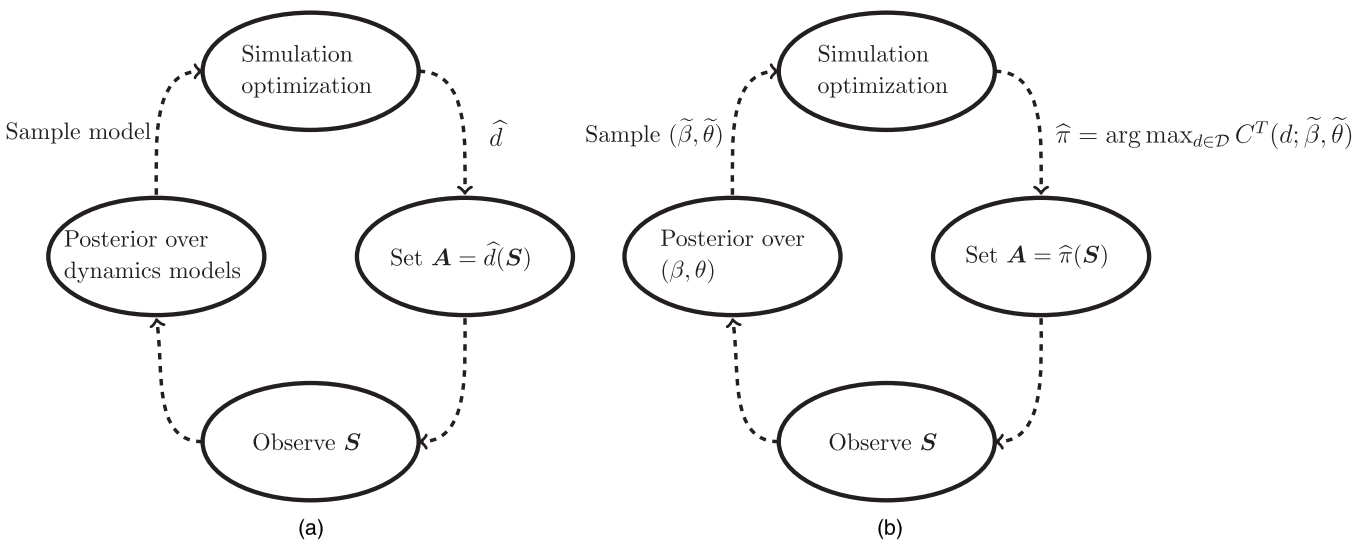

Fig. 2. (a) Schematic diagram for Thompson sampling over a generic class of dynamics models and (b) schematic diagram for Thompson sampling with a parametric class of models indexed by (θ, β) and a finite simulation horizon T

selected according to the posterior probability that they maximize the mean discounted cumulative utility (Thompson, 1933). Although Thompson sampling has been in print for more than 80 years it has only recently re-emerged in the computer science literature as a powerful tool for on-line decision making and has been shown to have several optimality properties in special cases (Chapelle and Li, 2011; Agrawal and Goyal, 2011; Kaufmann *et al.*, 2012; Korda *et al.*, 2013). Intuitively, a stochastic allocation strategy should balance exploration of the space of potential allocations with choosing allocations that are estimated to produce high expected utility; the proposed version of Thompson sampling achieves this balance through the posterior of mean utility under each $d \in \mathcal{D}$ which becomes increasingly concentrated on the maximizer as data accumulate.

To describe the implementation of our estimator we make several assumptions in addition to assumptions 1–3. We assume that the system is Markov and homogeneous in time so that, for any v, the densities in equation (1) become $f_v(\mathbf{y}^v|\mathbf{h}^v, \mathbf{a}^v) = f(\mathbf{y}^v|\mathbf{s}^v, \mathbf{a}^v)$ and $g_v(\mathbf{s}^v|\mathbf{h}^{v-1}) = g(\mathbf{s}^v|\mathbf{s}^{v-1}, \mathbf{a}^{v-1})$. Recall that the state $\mathbf{s}^v$ is a summary of the complete history up to time v, including past states, outcomes and allocations; see remark 1 for additional discussion. Furthermore, we assume parametric models $f(\mathbf{y}^v|\mathbf{s}^v, \mathbf{a}^v) = f(\mathbf{y}^v|\mathbf{s}^v, \mathbf{a}^v; \beta)$ and $g(\mathbf{s}^v|\mathbf{s}^{v-1}, \mathbf{a}^{v-1}) = g(\mathbf{s}^v|\mathbf{s}^{v-1}, \mathbf{a}^{v-1}; \theta)$ where β and θ are unknown parameters; under the model assumed, the system dynamics are completely determined by β and θ. If allocations are selected under the sequence of stochastic strategies $(\pi^1, \pi^2, \ldots, \pi^T)$, then the likelihood for (β, θ) given observed data $(\bar{\mathbf{S}}^T, \bar{\mathbf{Y}}^T, \bar{\mathbf{A}}^T)$ is

$$\mathfrak{L}_T(\beta, \theta) = \prod_{v=1}^{T} \{ f(\mathbf{Y}^v|\mathbf{S}^v, \mathbf{A}^v; \beta) \pi^v(\mathbf{A}^v; \mathbf{S}^v) g(\mathbf{S}^v|\mathbf{S}^{v-1}, \mathbf{A}^{v-1}; \theta) \},$$

where we define $g(\mathbf{s}^1|\mathbf{s}^0, \mathbf{a}^0) = g(\mathbf{s}^1)$ to be the distribution of the initial state.

For any deterministic strategy d and fixed $T > 0$, define

$$C^T(d; \beta, \theta) = \int \left\{ \sum_{t=1}^{T} \gamma^{t-1} u(\mathbf{y}^t) \right\} \prod_{v=1}^{T} [f\{\mathbf{y}^v|\mathbf{s}^v, d(\mathbf{s}^v); \beta\} g\{\mathbf{s}^v|\mathbf{s}^{v-1}, d(\mathbf{s}^{v-1}); \theta\}] \, \mathrm{d}\lambda(\bar{\mathbf{y}}^T, \bar{\mathbf{s}}^T),$$

and for $t \leqslant T$ define $\hat{\pi}^{t,T} = \arg\max_{d \in \mathcal{D}} C^T(d; \tilde{\beta}^t, \tilde{\theta}^t)$ where $\tilde{\beta}^t$ and $\tilde{\theta}^t$ are distributed according to the posterior of β, θ given $\mathbf{H}^t$. If the parametric densities are correctly specified, i.e. $f(\mathbf{y}^t|\mathbf{s}^t, \mathbf{a}^t) = f(\mathbf{y}^t|\mathbf{s}^t, \mathbf{a}^t; \beta^*)$ and $g(\mathbf{s}^t|\mathbf{s}^{t-1}, \mathbf{a}^{t-1}) = g(\mathbf{s}^t|\mathbf{s}^{t-1}, \mathbf{a}^{t-1}; \theta^*)$ for 'true' parameters β^* and θ^*, then, under standard regularity conditions (Gelman *et al.*, 2014), $\pi^{\mathrm{opt}} = \arg\max_{\pi \in \Pi} \lim_{t \to \infty} \{ \lim_{T \to \infty} C^T(d; \tilde{\beta}^t, \tilde{\theta}^t) \}$ with probability 1.

Algorithm 1 in Table 1 shows the procedure for estimating π^{opt} by using policy search with a system dynamics model and Thompson sampling; Fig. 2(b) displays a schematic diagram for this algorithm. The computational complexity of this algorithm depends on

Table 1. Algorithm 1: policy search algorithm for an optimal allocation strategy

Input: $T < \infty, \mathbf{S}^1$
draw $\tilde{\beta}^1, \tilde{\theta}^1$ from the prior
compute $\hat{\pi}^1 = \arg\max_{d \in \mathcal{D}} C^T(d; \tilde{\beta}^1, \tilde{\theta}^1)$ (via algorithm 2 in Section 4.1)
for $j \geqslant 1$ *do*
  apply allocation $\mathbf{A}^j = \hat{\pi}^j(\mathbf{S}^j)$; observe $\mathbf{Y}^j$ and $\mathbf{S}^{j+1}$
  draw $\tilde{\beta}^{j+1}$ and $\tilde{\theta}^{j+1}$ from the posterior of (β, θ) given $\mathbf{H}^{j+1}$
  compute $\hat{\pi}^{j+1} = \arg\max_{d \in \mathcal{D}} C^T(d; \tilde{\beta}^{j+1}, \tilde{\theta}^{j+1})$ (via algorithm 2 in Section 4.1)
end

(a) the class of strategies $\mathcal{D}$,
(b) the complexity of posterior distribution for β and θ, and
(c) the desired accuracy of the numerical integration that is used to compute $C^T(d;\beta,\theta)$.

In Section 4.1, we provide a class of strategies under which sampling from $\pi(\mathbf{a};\mathbf{s})$ scales linearly in the number of locations, L, making it feasible even when L is of the order of tens of thousands. In most ecological applications, the dimensions of β and θ are orders of magnitude smaller than L, e.g. the 'gravity model' for WNS (Maher *et al.*, 2012) is determined by 13 parameters; thus, integrating over the posterior of these parameters is typically not a computational bottleneck. As detailed in the next section, we use stochastic approximation to compute $\arg\max_{\pi\in\Pi} C^T(d;\beta,\theta)$; the number of Monte Carlo replicates in the numerical integration that is used to approximate $C^T(d;\beta,\theta)$ is generally smaller than L.

Remark 1. The Markov dynamics assumption that was used above is always trivially true if $\mathbf{S}^t = \mathbf{H}^t$ for all t (more formally, let $\mathbf{S}^t \in \mathbb{R}^\infty$ and define $\mathbf{S}^t = (t, \mathbf{H}^t, \mathbf{0})$, where $\mathbf{0}$ is the zero element in $\mathbb{R}^\infty$). However, this choice of state is rarely useful in large systems as the growing dimension makes modelling difficult. Thus, the Markov assumption can be viewed as an assumption about the ability of domain experts and analysts to construct a concise summary of the past that captures all salient features of the decision problem. One approach is to construct the state by concatenating information from the past k time points where k is dictated by domain knowledge or estimated from historical data. State construction for Markov decision processes is currently an active area of research (Mahadevan, 2009; Sugiyama, 2015).

Remark 2. The assumption of a low dimensional parametric model for the transition may seem overly restrictive in some settings. However, this can be relaxed through sieves (e.g. Newey (1997)) or Bayesian non-parametric models (Xu *et al.*, 2016; Ghosal and van der Vaart, 2017).

4.1. A scalable class of allocation strategies

The class of allocation strategies $\mathcal{D}$ has a large influence on the quality of the estimated optimal decision strategy and the computational complexity of algorithm 1. We propose a flexible but computationally efficient class of allocation strategies that is designed to scale to large decision problems with potentially tens of thousands of locations. However, as we demonstrate in the next section, this class of strategies is also useful for problems with as few as 100 locations. Throughout, we assume that at time t exactly c^t locations can be treated; although c^t is allowed to depend on the state $\mathbf{S}^t$ we suppress this in the notation.

The proposed class of allocation strategies is based on a parametric scoring function that assigns to each location a scalar priority score with high values indicating a greater need for treatment. Because of spatial interference, the priority score at a given location must take into account the state of that location, the states of nearby locations and the configuration of treatments at nearby locations. The optimal allocation for a given scoring function assigns treatment to the locations with highest priority scores with the number of treated locations dictated by resource constraints. Each value of the parameter vector indexing the priority score corresponds to a different scoring function and hence a different optimal allocation. However, for a given value of this parameter, computing the optimal allocation requires jointly optimizing over all allowable treatment allocations, which is computationally infeasible in all except the smallest problems. Instead, for each parameter indexing the priority score, we use a greedy batch updating algorithm to approximate the optimal allocation as follows. For simplicity, assume that resource constraints allow treatment of bm locations where m is the batch size and b is the number of batches (a formal description is given below). At the first stage of the batch optimization algo-

rithm the priority scores are calculated at each location by assuming that no other locations will be treated and the m locations with highest priority scores are selected to be treated. The priority scores are then recomputed at each location by assuming that those m locations selected at the first step will be treated and the m locations with highest priority scores (required to be distinct from those selected in the first step) are selected to be treated. Then, at each subsequent step the priority scores are updated assuming that those locations selected at previous steps will be treated and the m locations with highest priority scores are selected to be treated. This procedure is applied b times until there are a total of mb locations selected for treatment. After these mb locations have been selected the procedure either terminates or can continue iterating by selecting m of the selected mb locations to be set to untreated and recomputing the priority scores and subsequently selecting m new locations for treatment. The preceding procedure requires computing a parametric score at each location for each batch update; thus, the batch size dictates how computationally expensive the updates are. In applications, the batch size can be chosen to be large initially and reduced experimentally to see whether there is any change in the solution.

The class of allocation strategies that we propose depends on a parametric class of functions from $\mathrm{supp}(\mathbf{S})^t \times \{0,1\}^L$ into $\mathbb{R}^L$, $\mathcal{R} = \{R(\mathbf{s}^t, \mathbf{a}^t; \eta) : \eta \in E\}$, where $E \subseteq \mathbb{R}^q$. Given $\eta \in E$, the function $R(\mathbf{s}^t, \mathbf{a}^t; \eta)$ is a vector of priority scores, one per location, so $R_l(\mathbf{s}^t, \mathbf{a}^t; \eta)$ represents the priority for treating location l at time t if the observed state is $\mathbf{S}^t = \mathbf{s}^t$ and assuming that the locations $\{j : a_j^t = 1\}$ are certain to be treated. If $a_l^t = 1$ then $R_l(\mathbf{s}^t, \mathbf{a}^t; \eta) = -\infty$ so each location is selected for treatment at most once per time point. For each non-negative integer m, define the binary vector

$$\mathcal{U}_l^t(\mathbf{s}^t, \mathbf{a}^t; \eta, m) = \begin{cases} 1 & \text{if } R_l(\mathbf{s}^t, \mathbf{a}^t; \eta) \geqslant R_{(m)}(\mathbf{s}^t, \mathbf{a}^t; \eta), \\ 0 & \text{otherwise,} \end{cases}$$

where $l \in \mathcal{L}$ and $R_{(k)}(\mathbf{s}^t, \mathbf{a}^t; \eta)$ denotes the kth order statistic of $\{R_l(\mathbf{s}^t, \mathbf{a}^t; \eta)\}_{l \in \mathcal{L}}$. Let $k \leqslant c^t$ be a non-negative integer and $\mathbf{0}$ to be a vector of 0s. Define $d^{(1)}(\mathbf{s}^t; \eta)$ to be the binary vector that selects the $\lfloor c^t/k \rfloor$ locations with the highest priority scores. Let $w^{(1)}$ denote $d^{(1)}(\mathbf{s}^t; \eta)$. Recursively, for $j = 2, \ldots, k$, set $w^{(j)} = d^{(j)}(\mathbf{s}^t; \eta)$, $\Delta_j = \lfloor jc^t/k \rfloor - \lfloor (j-1)c^t/k \rfloor$ and $d^{(j)}(\mathbf{s}^t, \eta) = \mathcal{U}^t\{\mathbf{s}^t, w^{(j-1)}; \eta, \Delta_j\} + w^{(j-1)}$. The final decision rule is $d(\mathbf{s}^t; \eta) = d^{(k)}(\mathbf{s}^t; \eta)$; the dependence of this rule on t occurs only through c^t.

The parameter k in the above class of strategies governs the number of locations that are selected each time that the priority scores are updated. If $k = 1$ then the priority scores are computed once, under no treatments, and the top c^t locations are treated; if $k = c^t$ then the algorithm updates the priority scores after every location selection. In large problems, we anticipate choosing $k \ll L$, e.g. $k = O\{\log(L)\}$. If the computational complexity of computing $R_l(\mathbf{s}^t, \mathbf{a}^t; \eta)$ is N, then the complexity of computing $d(\mathbf{s}^t; \eta)$ is $O(kLN)$. Thus, if $k = O\{\log(L)\}$ and N is negligible relative to L, then evaluating the strategy is $O\{L \log(L)\}$ which is feasible even for large values of L.

Let $\mathcal{D}$ denote the class of policies $\{d(\mathbf{s}; \eta) : \eta \in E\}$. Algorithm 1 requires maximization of $C^T(d; \beta, \theta)$ over $d \in \mathcal{D}$ (or, equivalently, over $\eta \in E$). Thus, the order of computation for this step depends on the complexity of evaluating the function $C^T(d; \beta, \theta)$. Obtaining a high quality approximation for $C^T(d; \beta, \theta)$ for any fixed strategy d may require a large number of expensive Monte Carlo replications; this is particularly wasteful when evaluating allocation strategies that are far from a maximizer. Thus, we use a stochastic approximation algorithm which is known as simultaneous perturbation (e.g. Spall (2005)) to approximate $\arg\max_{d \in \mathcal{D}} C^T(d; \beta, \theta)$. The algorithm relies on a sequence of non-negative step sizes $\{\alpha_j\}_{j \geqslant 1}$ and a sequence of non-negative $\{\zeta_j\}_{j \geqslant 1}$ that satisfies $\zeta_j \to 0$ as $j \to \infty$. Convergence guarantees require that $\Sigma_{j \geqslant 1} \alpha_j = \infty$, $\Sigma_{j \geqslant 1} \alpha_j^2 < \infty$ and $\zeta_j \to 0$ as $j \to \infty$ (Kushner and Yin, 2003; Borkar, 2008); however, In our

Table 2. Algorithm 2: stochastic approximation

Input: $T<\infty$, $\mathbf{S}^t$, $\eta^0 \in E$, $f(\mathbf{y}^t|\mathbf{s}^t,\mathbf{a}^{t-1};\beta)$, $g(\mathbf{s}^t|\mathbf{s}^{t-1},\mathbf{a}^t;\theta)$, $\{\alpha_j\}_{j\geqslant 1}$, $\{\zeta_j\}_{j\geqslant 1}$ and tol >0

set $k=1$, $\tilde{\mathbf{S}}^t=\mathbf{S}^t$
do
  draw $\mathbf{Z}^k \sim \text{uniform}\{-1,1\}^d$
  for $m=0,\ldots,T-1$ *do*
    set $\mathbf{A}^{t+m}=d(\mathbf{S}^{t+m};\eta^k+\zeta_k\mathbf{Z}^k)$
    draw $\mathbf{S}^{t+m+1}\sim g(\mathbf{s}^{t+m+1}|\mathbf{S}^{t+m},\mathbf{A}^{t+m};\theta)$
    draw $\mathbf{Y}^{t+m}\sim f(\mathbf{y}^{t+m}|\mathbf{S}^{t+m},\mathbf{A}^{t+m};\beta)$
    set $\tilde{\mathbf{A}}^{t+m}=d(\tilde{\mathbf{S}}^{t+m};\eta^k-\zeta_k\mathbf{Z}^k)$
    draw $\tilde{\mathbf{Y}}^{t+m}\sim f(\mathbf{y}^{t+m}|\tilde{\mathbf{S}}^{t+m},\tilde{\mathbf{A}}^{t+m};\beta)$
    draw $\tilde{\mathbf{S}}^{t+m+1}\sim g(\mathbf{s}^{t+m+1}|\tilde{\mathbf{S}}^{t+m},\tilde{\mathbf{A}}^{t+m};\theta)$
  end
  set $\eta^{k+1}=\mathbb{G}_E[\eta^k+(\alpha_k/2\zeta_k)\{\mathbf{Z}^k\mathbf{1}_L^{\mathrm{T}}(\mathbf{Y}^{t+T-1}-\tilde{\mathbf{Y}}^{t+T-1})\}]$
  set $k=k+1$
while $\alpha_k \geqslant$ tol
Output: η^k

simulation experiments, we use $\alpha_j=\tau/(\rho+j)^{1.25}$ and $\zeta_j=100/j$, where $\tau,\rho>0$ are tuning parameters. We used a double-bootstrap procedure to select τ and ρ; details on the double-bootstrap tuning procedure are in the on-line supplemental materials. Let $\mathbb{G}_E$ denote the orthogonal projection onto E, i.e. $\mathbb{G}_E(\rho)=\arg\min_{\eta\in E}||\rho-\eta||$, where $||\cdot||$ denotes the Euclidian norm. Algorithm 2 in Table 2 shows the stochastic approximation algorithm for computing $\arg\max_{d\in\mathcal{D}} C^T(d;\beta,\theta)$; it can be seen that each iteration of this algorithm requires simulating trajectories under only two parameter values, rather than $O(d)$ parameter values as would be required by classic stochastic gradient descent methods using a difference-based approximation for the gradient (Spall, 2005; Bhatnagar *et al*., 2013).

Remark 3. The choice of priority functions $\mathcal{R}$ will, of course, depend on features of a given application. However, for concreteness, we describe a class of linear priority functions that may be useful in practice. We use linear priority functions in our application to WNS. For each $l\in\mathcal{L}$ let $\phi_l(\mathbf{s}^t,\mathbf{a}^t)\in\mathbb{R}^p$ denote a fixed and known feature vector. Linear priority functions $\mathcal{R}_{\text{Lin}}=\{R_l(\mathbf{s}^t,\mathbf{a}^t;\eta)=\phi_l(\mathbf{s}^t,\mathbf{a}^t)^{\mathrm{T}}\eta : l\in\mathcal{L},\eta\in\mathbb{R}^p,||\eta||=1\}$, are appealing because the priority scores are interpretable and computationally simple. The features $\phi_l(\mathbf{s}^t,\mathbf{a}^t)$ can be local smooths of location covariates, e.g. measures of susceptibility or contagiousness, predictions of the disease process based on one or more postulated models for the underlying system dynamics or structural characteristics (Kolaczyk, 2009).

5. Simulation experiments

We evaluate the finite sample performance of our estimator in a suite of simulation experiments. We consider the spread of an infectious disease over two spatial domains:

(a) points in two-dimensional Euclidean space and
(b) nodes in a network.

5.1. Spread of an infectious disease in Euclidean space

In the Euclidean space setting, each location $l\in\{1,\ldots,L\}$ is a point in the unit square $[0,1]^2$.

For each location l, we generate four static covariates $\mathbf{X}_l$ by using a mean 0 Gaussian process with a multivariate separable isotropic covariance matrix that is exponential in space and autoregressive across the four covariates at each location. To mimic our motivating example of the spread of WNS, we assume that each location has a fixed number of caves; in more general spatial epidemic models, this variable represents the gravity that is associated with the location (e.g. Bossenbroek *et al.* (2001), Xia *et al.* (2004), Drake and Lodge (2004) and Sen and Smith (2012)). We generate the number of caves, Z_l, by using the first covariate and subtracting the minimum value to force non-negative values (see the on-line supplemental materials for details). The outcome Y_l^t is 1 if location l becomes infected at or before time t and 0 otherwise. The process is initialized at time $t=1$ by randomly selecting 1% of the locations to be infected. Define $\mathbf{S}_l^t = (\mathbf{X}_l, Z_l, Y_l^{t-1})$ and let $\omega_{l,k}$ denote the distance between locations l and k. We standardize the distance to have a standard deviation of 1 for computational stability. The model assumes that disease transmission is independent across locations given $(\mathbf{S}^t, \mathbf{A}^t)$, with $Y_l^t = 1$ if $Y_l^{t-1} = 1$ so that locations never shed the disease, and $P(Y_l^t = 1 | \mathbf{S}^t = \mathbf{s}^t, \mathbf{A}^t = \mathbf{a}^t) = 1 - \Pi_{k \in \mathcal{I}^t}\{1 - q_{l,k}(\mathbf{s}^t, \mathbf{a}^t)\}$ if $Y_l^{t-1} = 0$, where $\mathcal{I}^t = \{k : Y_k^{t-1} = 1\}$ is the set of sites that are infected before time t, and $q_{l,k}(\mathbf{s}^t, \mathbf{a}^t)$ is the probability that the disease spreads from location k to location l. The local dynamics are governed by the following spatial gravity model (Maher *et al.*, 2012):

$$\text{logit}\{q_{l,k}(\mathbf{s}^t, \mathbf{a}^t)\} = \theta_0 + \mathbf{x}_l^{\mathrm{T}}\boldsymbol{\theta}_1 + \mathbf{x}_k^{\mathrm{T}}\boldsymbol{\theta}_2 - \theta_3 a_l^t - \theta_4 a_k^t - \theta_5 \omega_{lk}/(z_l z_k)^{\theta_6}, \tag{2}$$

where θ_0 is an intercept, $\boldsymbol{\theta}_1$ captures effects of the uninfected location, $\boldsymbol{\theta}_2$ captures effects of the infected location, θ_3 and θ_4 govern the strength of treatments to uninfected and infected locations, $\theta_5 > 0$ controls the spatial range of infection and $\theta_6 > 0$ controls the amount of gravity that is induced by the number of caves per location. We chose $\boldsymbol{\theta} = (\theta_0, \boldsymbol{\theta}_1^{\mathrm{T}}, \boldsymbol{\theta}_2^{\mathrm{T}}, \theta_3, \ldots, \theta_6)^{\mathrm{T}}$ to match the maximum likelihood estimator fit using data on WNS but adjusted θ_0 so that in each simulation setting an average 70% of locations become infected after $T = 15$ years in the absence of any interventions. An algorithm for constructing $\boldsymbol{\theta}$ for each simulation setting is provided in the on-line supplemental materials.

We assume that treatments can be applied to at most only ςL locations at each time t where $\varsigma \in (0, 1)$ (in all settings considered, ςL is an integer). To form a baseline for comparison, we also consider the following allocation strategies:

(a) no treatment, do not apply treatment at any location;
(b) myopic, rank uninfected locations by their estimated probability of becoming infected in the next time step, rank infected locations by the weighted average infection probability of uninfected locations by using $\lambda_{l,k}$ (defined below) as weights and then allocate treatment to the $(\varsigma/2)L$ highest ranked uninfected locations and to the $(\varsigma/2)L$ highest ranked infected locations;
(c) proximal, rank uninfected locations by their proximity (inverse distance) to the nearest infected location and rank infected locations by their proximity to uninfected locations, allocate treatment to the $(\varsigma/2)L$ highest ranked infected locations and to the $(\varsigma/2)L$ highest ranked uninfected locations;
(d) treat, *all* locations to provide a performance ceiling if infinite resources were available and, by contrast with the no-treatment strategy, to provide a sense of the strength of treatment.

In the simulations that we present here, we set $\varsigma = 0.12$; additional simulations with other settings of ς are qualitatively similar and have thus been omitted. Because the number of possible treatment allocations is exponential in the number of locations, it is not computationally feasible to compute the optimal allocation strategy as a 'gold standard' even though the generative

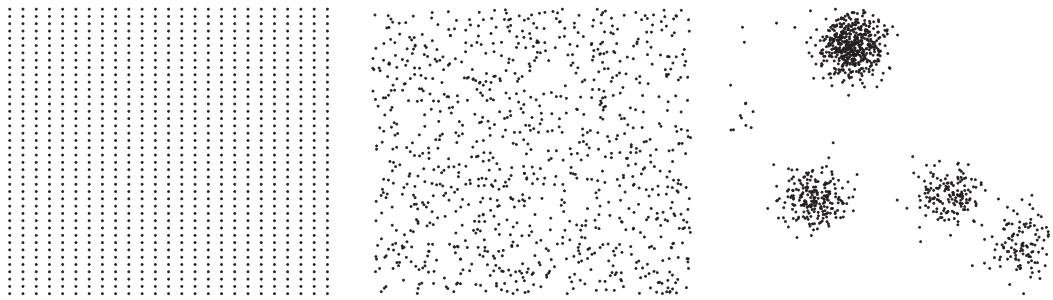

Fig. 3. (a) S1, regular lattice layout with 1000 locations, (b) S2, uniformly distributed layout with 1000 locations, and (c) S3, clustered layout with 1000 locations.

distribution is known; for example in the smallest setting we consider there are $\binom{100}{12} \approx 10^{15}$ possible allocations at each time point.

We consider three spatial location layouts: S1, a regular lattice layout, S2, a uniformly distributed layout, and S3 a clustered layout. Instances of these three layouts with $L = 1000$ locations are displayed in Fig. 3. Our simulation experiments consider location sets of size $L = 100, 500, 1000$; algorithms for generating these layouts are in the on-line supplemental Materials. The infection is allowed to spread from $t = 1$ to $t = 8$ with no interventions. At time points $t = 8, \ldots, 15$, each strategy under consideration is used to choose a treatment allocation. Performance of the allocation strategies is measured in terms of the average proportion infected after $T = 15$ time points. Because there is a fixed and finite time horizon, in the implementation of our algorithm we set $\gamma = 1$.

To study the effects of model misspecification, we also consider the case when the true data-generating model is model (2) and the allocation strategies proposed are based on the postulated model (2) with $\text{logit}\{q^t_{l,k}(\mathbf{s}, \mathbf{a})\}$ replaced with $\tilde{\theta}_0 + \tilde{\theta}_1 \omega_{j,l} - \tilde{\theta}_2 a^t_l - \tilde{\theta}_3 a^t_j$ if site l is uninfected at time t and 0 otherwise. Thus, the misspecified model assumes that under no treatment the probability of spread from an infected to susceptible location is dictated by distance only.

We use linear priority functions as in remark 3: $\mathcal{R}_{\text{Lin}} = \{R_l(\mathbf{s}^t, \mathbf{a}^t; \eta) = \phi_l(\mathbf{s}^t, \mathbf{a}^t)^{\text{T}} \eta : l \in \mathcal{L}, \eta \in \mathbb{R}^p, ||\eta|| = 1\}$. Let $m^t_{l,k}(\mathbf{s}^t, \mathbf{a}^t)$ denote the, possibly misspecified, postulated model for $q^t_{l,k}(\mathbf{s}^t, \mathbf{a}^t)$; write P_m to denote probabilities that are evaluated under the model postulated. Let c_l denote the half-plane data depth of location l (i.e. the minimum number of points that are contained in a hyperlane passing through l; Liu (1990)) and define $\lambda_{l,j} = \exp(\lambda \omega_{l,j}) / \Sigma_{i \neq l} \exp(\lambda \omega_{l,j})$ where the constant λ is chosen so that 80% of the total weight is placed on the $\log(L)$ nearest neighbours of location l. Recall that $\mathcal{I}^t$ denotes the set of locations that were infected before time t; let $\tilde{\mathcal{I}}^t$ denote the complement of this set. For uninfected locations, at time t, define

$$\psi_{l,1}(\mathbf{s}^t, \mathbf{a}^t) = P_m(Y^t_l = 1 | \mathbf{S}^t = \mathbf{s}^t, \mathbf{A}^t = \mathbf{a}^t),$$
$$\psi_{l,2}(\mathbf{s}^t, \mathbf{a}^t) = \psi_{l,1}(\mathbf{s}^t, \mathbf{a}^t) \sum_{j \in \mathcal{I}^t} \{1 - \psi_{j,1}(\mathbf{s}^t, \mathbf{a}^t)\} m_{l,j}(\mathbf{s}^t, \mathbf{a}^t) \lambda_{l,j},$$
$$\psi_{l,3}(\mathbf{s}^t, \mathbf{a}^t) = c_l \psi^t_{l,1}(\mathbf{s}, \mathbf{a}).$$

Thus, for uninfected locations, $\psi_{l,1}(\mathbf{s}^t, \mathbf{a}^t)$ is the probability that location l becomes infected at the next time point; $\psi_{l,2}(\mathbf{s}^t, \mathbf{a}^t)$ is the effect of treating location l on the infection probabilities of all other locations. The third feature, $\psi_{l,3}(\mathbf{s}^t, \mathbf{a}^t)$, is the probability that location l becomes infected at the next time point weighted by the data depth of location l. For infected locations, define

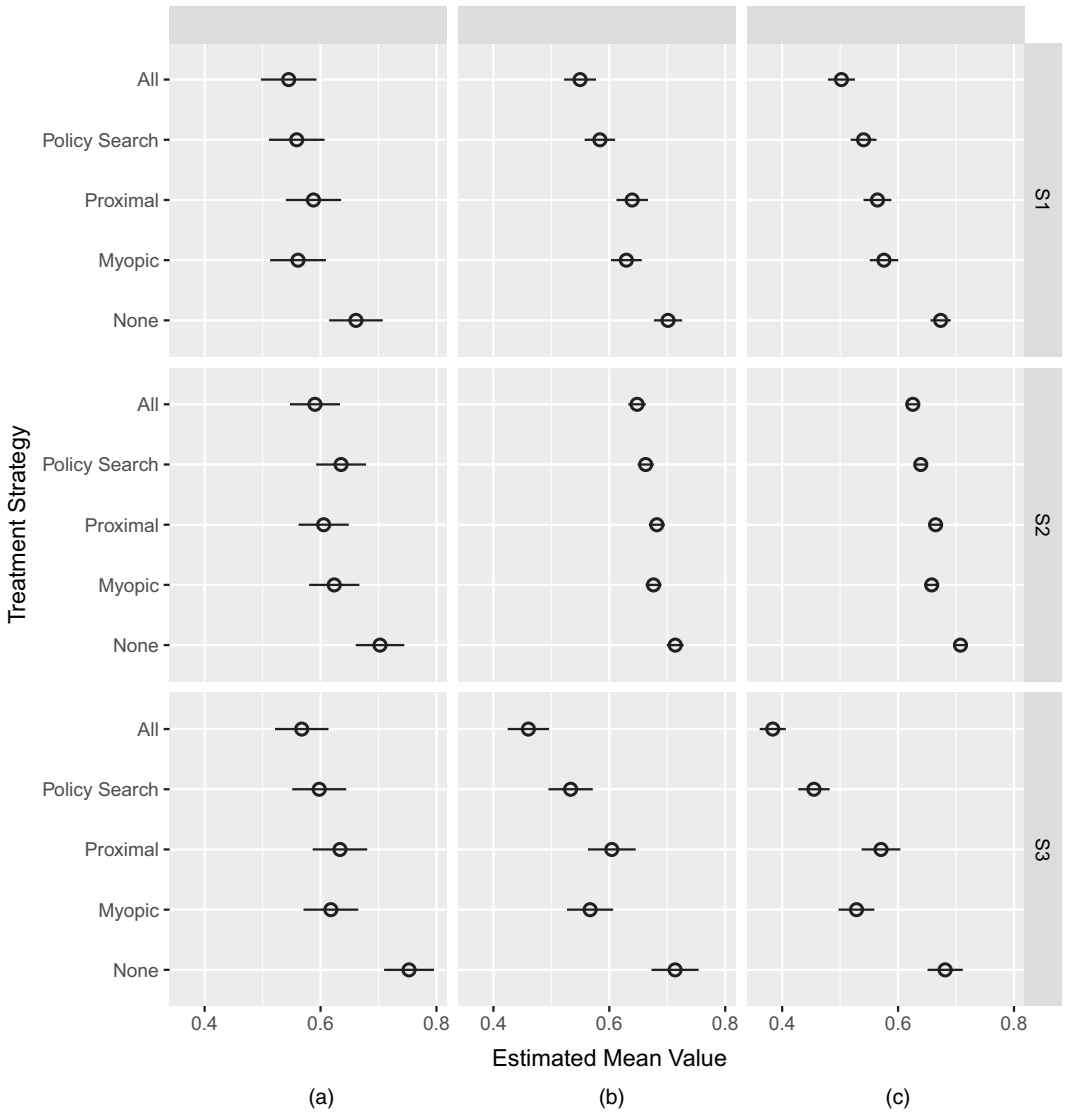

Fig. 4. Estimated average proportion infected based on 100 Monte Carlo replications under correct specification of the system dynamics model (—, 2 standard errors): (a) $L = 100$; (b) $L = 500$; (c) $L = 1000$

$$\phi_{l,1}(\mathbf{s}^t, \mathbf{a}^t) = \sum_{j \in \tilde{\mathcal{I}}^t} \lambda_{l,j} \psi_{j,1}(\mathbf{s}^t, \mathbf{a}^t),$$
$$\phi_{l,2}(\mathbf{s}^t, \mathbf{a}^t) = \sum_{j \in \tilde{\mathcal{I}}^t} \psi_{j,2}(\mathbf{s}^t, \mathbf{a}^t) m^t_{l,j}(\mathbf{s}^t, \mathbf{a}^t),$$
$$\phi_{l,3}(\mathbf{s}^t, \mathbf{a}^t) = \sum_{j \in \tilde{\mathcal{I}}^t} c_j m^t_{l,j}(\mathbf{s}^t, \mathbf{a}^t).$$

Thus, for infected locations, $\phi_{l,1}(\mathbf{s}^t, \mathbf{a}^t)$ is the weighted average infection probability over all uninfected locations at time t; $\phi_{l,2}(\mathbf{s}^t, \mathbf{a}^t)$ is a measure of expected secondary infections stemming from location l; $\phi_{l,3}(\mathbf{s}^t, \mathbf{a}^t)$ is data depth weighted by probability of infection by l.

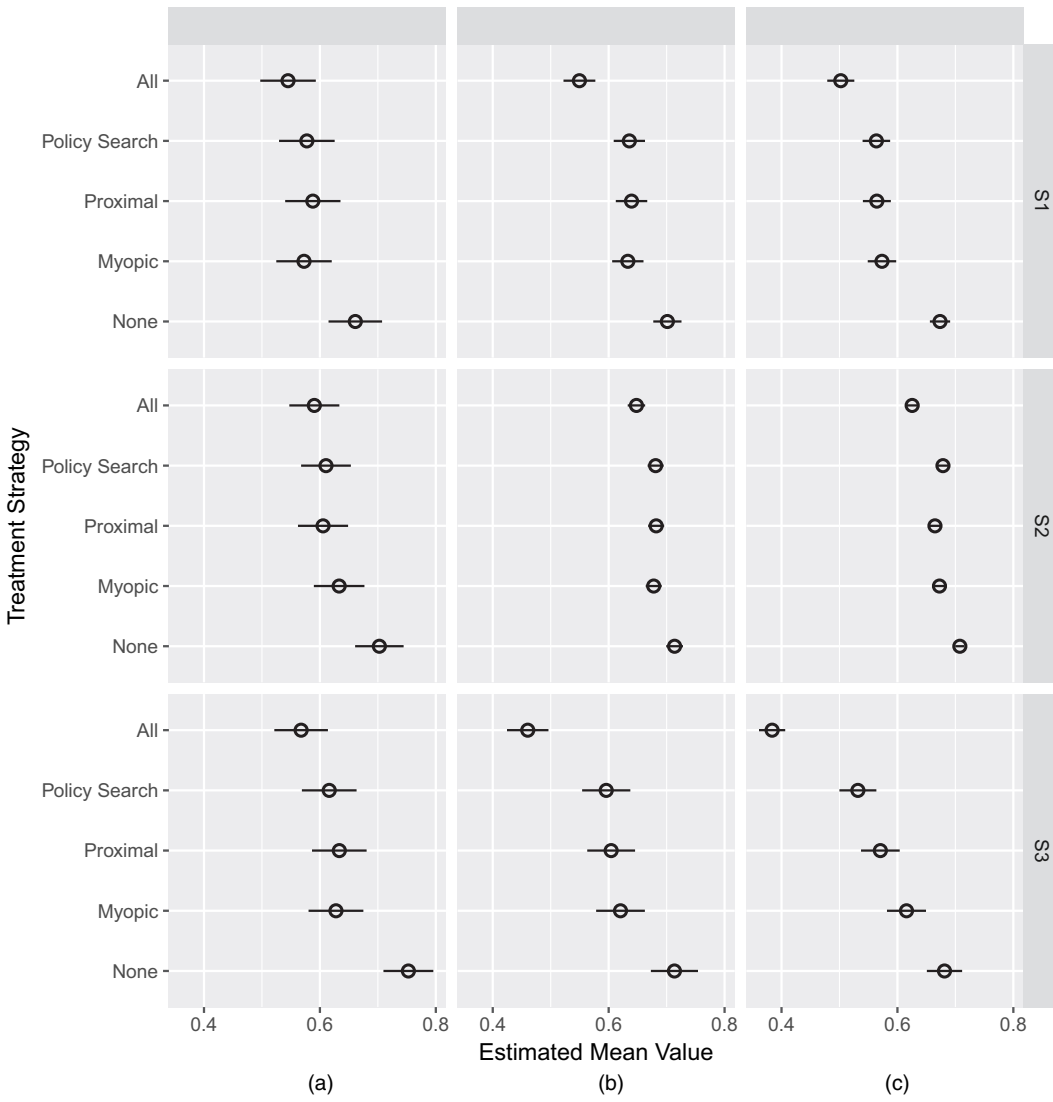

Fig. 5. Estimated average proportion infected based on 100 Monte Carlo replications under misspecification of the system dynamics model (—, 2 standard errors): (a) $L = 100$; (b) $L = 500$; (c) $L = 1000$

To reduce the computational burden, we approximate the posterior distribution of $\boldsymbol{\theta}$ by using the plug-in estimator of the sampling distribution of the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n$; however, at the first intervention period, the treatment effects are not estimable so we sample them from the prior distribution. Additional simulations (which are not shown here) suggested that this approximation reduced the computation time by two orders of magnitude and did not affect the quality of the solution.

Simulation results under a correctly specified dynamics model are presented in Fig. 4. In all settings, the policy search algorithm proposed resulted in a smaller average proportion of infected locations than those of competing methods. Results for the misspecified dynamics model are presented in Fig. 5. As expected, the performance of policy search is worse under

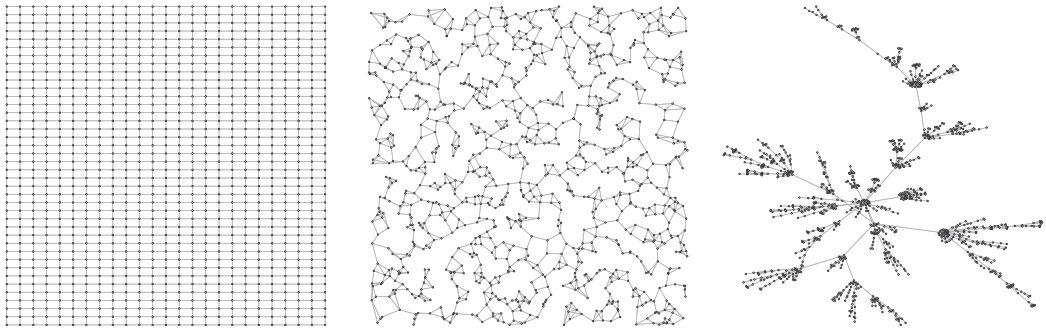

Fig. 6. (a) N1, a regular lattice network with 1000 locations, (b) N2, a random k-nearest-neighbour network with 1000 locations, and (c) N4, a small world network with 1000 locations

the incorrectly specified dynamics model; however, it still performed favourably relatively to competing methods.

5.2. Spread of an infectious disease across a network

In the network setting, each location represents a node in a network with an adjacency matrix $\mathbf{\Omega}$, i.e. $\Omega_{l,k} = 1$ if location l and k are adjacent and $\Omega_{l,k} = 0$ otherwise. Define $\mathcal{N}_l = \{k : \Omega_{l,k} = 1\}$ to be the set of adjacent locations to location l. We assume that the infection can only spread along edges in the network. Thus, if uninfected location l has zero infected neighbours at time t, so that $\mathcal{N}_l \cap \mathcal{I}_t = \varnothing$, then location l will remain uninfected at time $t+1$ with probability 1. The probability of infection is $P(Y_l^t = 1 | \mathbf{S}^t = \mathbf{s}^t, \mathbf{A}^t = \mathbf{a}^t) = 1 - \Pi_{k \in \mathcal{N}_l \cap \mathcal{I}^t}\{1 - q_{l,k}(\mathbf{s}^t, \mathbf{a}^t)\}$ where the product is now over only infected locations adjacent to location l. We define $\Pi_{k \in \varnothing} =^{\triangle} 1$ so that the probability of infection for an uninfected node is 0 when none of its neighbours are infected. The distance between any two locations is defined as the number of edges along the shortest path between them. Thus, we set $\theta_5 = 0$ in the gravity model as only neighbours of an uninfected location can influence its probability of infection. The generative model is

$$\text{logit}\{q_{l,k}(\mathbf{s}^t, \mathbf{a}^t)\} = \theta_0 + \mathbf{x}_l^{\mathrm{T}} \boldsymbol{\theta}_1 + \mathbf{x}_k^{\mathrm{T}} \boldsymbol{\theta}_2 - \theta_3 a_l^t - \theta_4 a_k^t. \tag{3}$$

As in the preceding section, we chose $\boldsymbol{\theta}$ to mimic the spread of WNS and adjusted the intercept θ_0 so that 70% of locations will be infected on average after $T = 15$ if no interventions are applied. Parameter values that were used in the simulation and details about fitting a network spread model to the observed WNS data can be found in the on-line supplementary materials. For misspecification of the system dynamics model, we remove covariate information, leaving an intercept and treatment effects. The misspecified model has the form

$$\text{logit}\{q_{l,k}(\mathbf{s}^t, \mathbf{a}^t)\} = \tilde{\theta}_0 - \tilde{\theta}_1 a_l^t - \tilde{\theta}_2 a_k^t. \tag{4}$$

We consider the following network structures: N1, a lattice, N2, a random three-nearest neighbour graph, and N3 a small world network. Instances of these networks are displayed in Fig. 6. We use the same class of treatment strategies based on linear priority scores as in the preceding section. However, we redefine $\psi_{l,3}(\mathbf{s}^t, \mathbf{a}^t)$ and $\phi_{l,3}(\mathbf{s}^t, \mathbf{a}^t)$ to reflect distance as measured along paths in the netwrok. In the spatial setting, we set c_l to the half-plane data depth of location l; as locations are now nodes in a network, we set c_l to be the subgraph centrality of location l (Estrada and Rodriguez-Velazquez, 2005). Additionally, because infection can spread only between adjacent nodes, we set $\lambda_{l,k} = \Omega_{l,k}$ for all l and k.

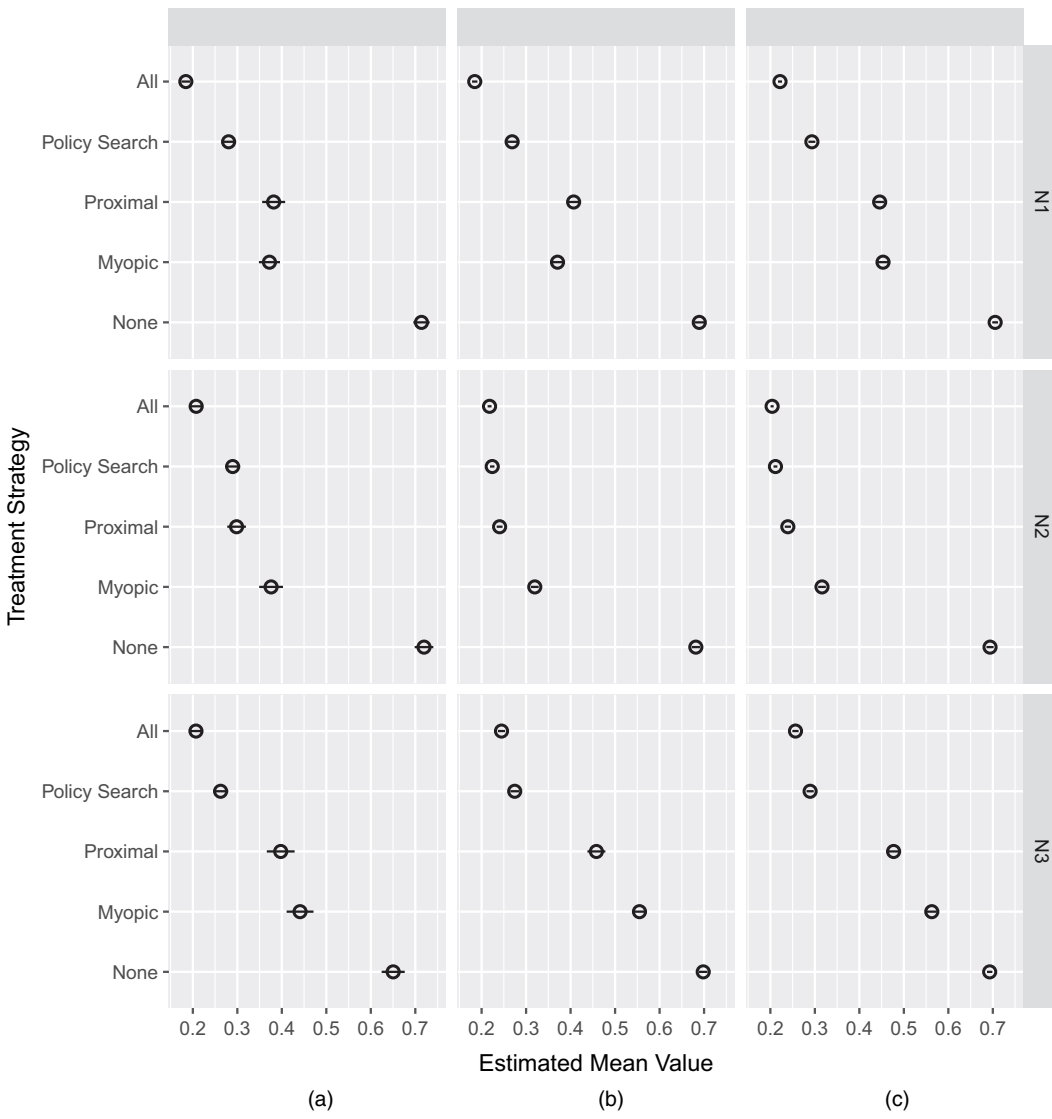

Fig. 7. Estimated average proportion infected based on 100 Monte Carlo replications under correct specification of the network spread dynamics model (—, 2 standard errors): (a) $L = 100$; (b) $L = 500$; (c) $L = 1000$

Our simulations for the spread over a network use the same competing methods and performance measures as the spread over Euclidean space. Fig. 7 shows the result when the dynamics model is correctly specified and Fig. 8 shows the results when the model is incorrectly specified. As in the preceding section, policy search performs favourably to competitors even when the dynamics model is misspecified.

6. Controlling the spread of white nose syndrome

One motivation for this work is the need to design a treatment allocation strategy to inform the management of WNS. Fig. 1 shows the current (at the time of writing) reported spread of

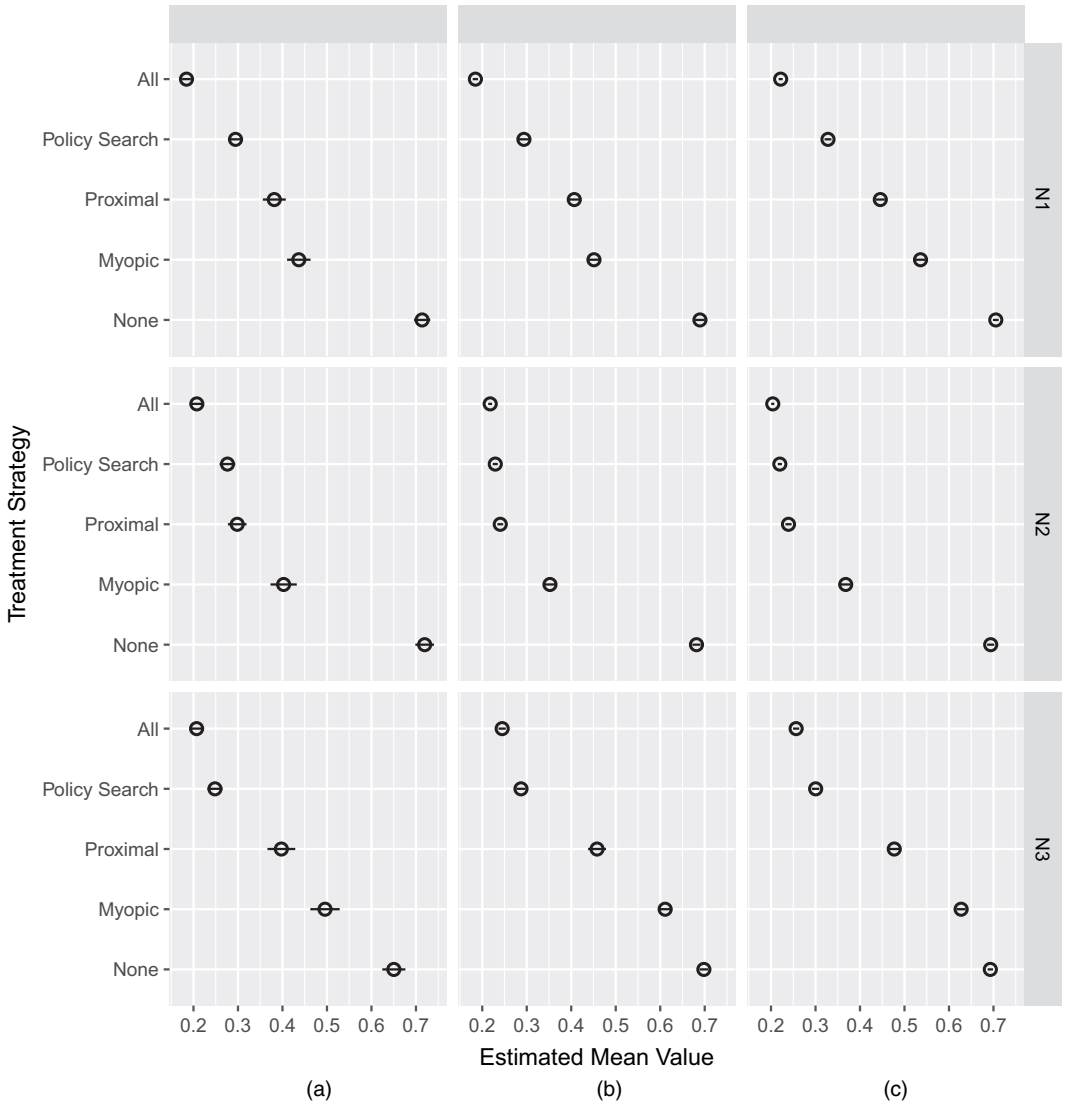

Fig. 8. Estimated average proportion infected based on 100 Monte Carlo replications under misspecification of the network spread dynamics model (—, 2 standard errors): (a) $L = 100$; (b) $L = 500$; (c) $L = 1000$

WNS. The locations in this example are cave bearing counties that are known to house bats that are susceptible to WNS. This disease is still emerging; thus the nature of the contagion and potential treatments are still under study (e.g. Field *et al*. (2014), O'Donoghue *et al*. (2015), Turner *et al*. (2015) and Bernard *et al*. (2015)). Our goal is to use existing data on the spread of WNS to construct an allocation strategy that could be deployed as soon as viable treatments are available. We evaluate the performance of the estimated allocation strategy by simulating the spread of WNS from 2015 to 2022 under a postulated system dynamics model.

6.1. A system dynamics model for white nose syndrome under no interventions

We begin by fitting the gravity model (2) to the WNS data that are plotted in Fig. 1. We use

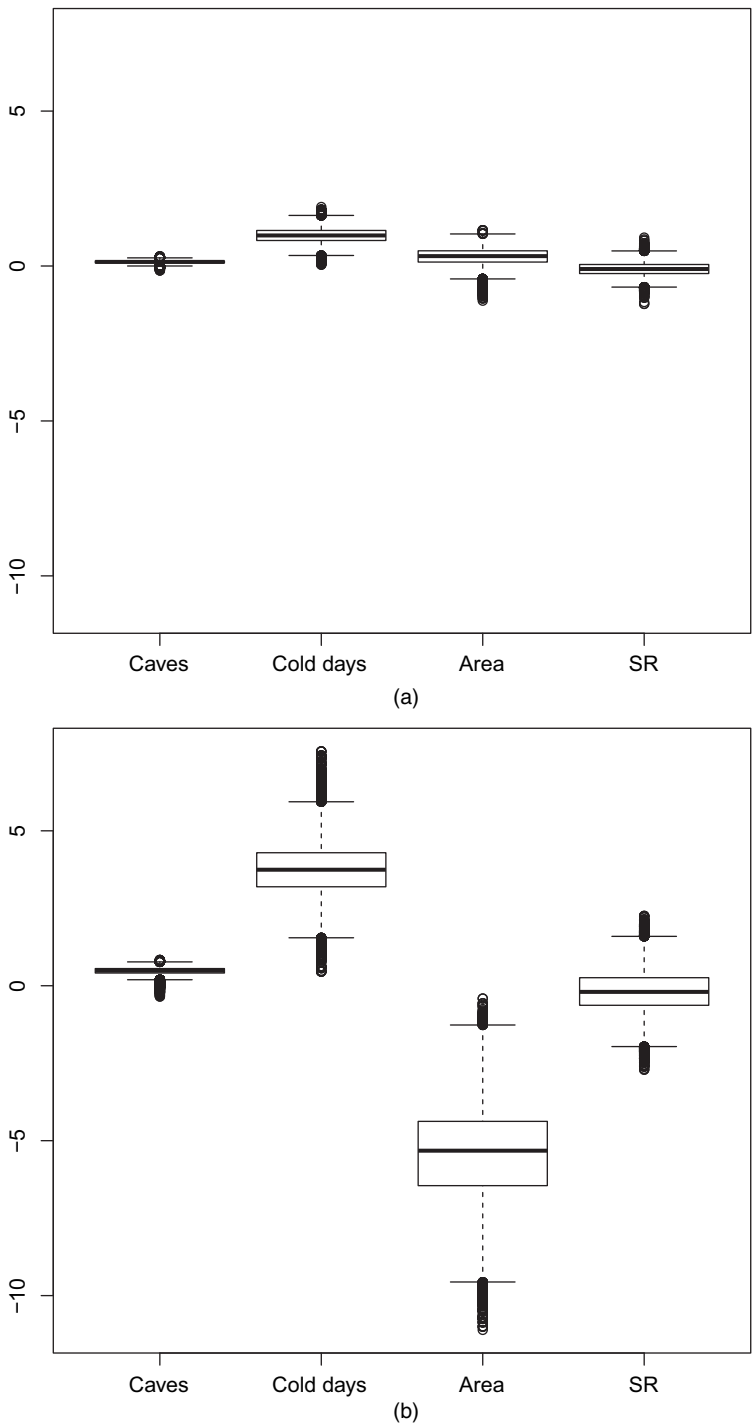

Fig. 9. Posterior distribution of the regression parameters associated with (a) uninfected (θ_1) and (b) infected (θ_2) counties in the gravity model (2) applied to the WNS data: the covariates are the number of caves in the county ('caves'), the average number of days per year below 10 °C ('cold days'), area in kilometres squared ('Area') and species richness ('SR')

Table 3. Estimated coefficients and 95% credible intervals (CIs) using for the gravity model fit using WNS data from 2006–2014†

Parameter	*Posterior mean*	*95% credible interval*	*Value used in simulation*
θ_0	−8.35	[−10.42, −6.62]	−8.35
$\theta_{1,0}$	0.13	[0.03, 0.22]	0.13
$\theta_{1,1}$	0.99	[0.51, 1.45]	0.99
$\theta_{1,2}$	0.30	[−0.31, 0.77]	0.30
$\theta_{1,3}$	−0.10	[−0.53, 0.33]	−0.10
$\theta_{2,0}$	0.48	[0.23, 0.69]	0.48
$\theta_{2,1}$	3.76	[2.18, 5.44]	3.76
$\theta_{2,2}$	−5.45	[−8.66, −2.67]	−5.45
$\theta_{2,3}$	−0.19	[−1.53, 1.13]	−0.19
θ_3	—†	—‡, —‡	1.77
θ_4	—‡	—‡, —‡	1.77
θ_5	24.21	[15.56, 35.86]	24.21
θ_6	0.22	[0.17, 0.28]	0.22

†Rows for θ_3 and θ_4 correspond to intervention effects which are not identifiable in the WNS data as these data do not contain any interventions.
‡Not applicable.

Table 4. Average proportion of infected counties in 100 Monte Carlo simulations of the spread of WNS from 2015–2022 under the gravity model†

No treatment	*Proximal*	*Myopic*	*Treat all*	*Policy search*
0.43 (0.006)	0.39 (0.006)	0.36 (0.005)	0.17 (0.001)	0.22 (0.002)

†Policy search resulted in markedly fewer infected counties than did the next best estimator.

four static (centred and scaled) covariates in $\mathbf{X}_l$: the number of caves in the county 'Z_l' the average number of days per year with temperature below 10 °C; area (in square kilometres) and species richness (the number of bat species that are thought to occupy the county). No treatments have been applied and thus $A_l^t = 0$ for all l and t, and the distance $\omega_{k,l}$ is measured as kilometres between county centroids. We assume an $N(0, 10^2)$ prior for θ_0 and independent $N(0, 10)$ priors for the elements of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, and standard independent normal priors for $\log(\theta_5)$ and $\log(\theta_6)$. We sample from the posterior by using Metropolis sampling with Gaussian candidate distributions tuned to give an acceptance probability around 0.4; 100 000 samples are generated and the first 20 000 are discarded as burn-in.

The posterior is summarized in Fig. 9. As expected, the transmission probability is high when the infected and uninfected counties have many caves and many days below 10 °C. These factors increase the space and time for hibernation, and it is believed that the disease spreads primarily via contact between bats hibernating in a cave. Also, the transmission probability is high when the infected county is small, presumably because the disease can rapidly spread through the infected county and thus move quickly to nearby counties. The on-line supplemental materials include model comparisons and posterior predictive model checks which suggest that this relatively simple model is adequate for our purposes.

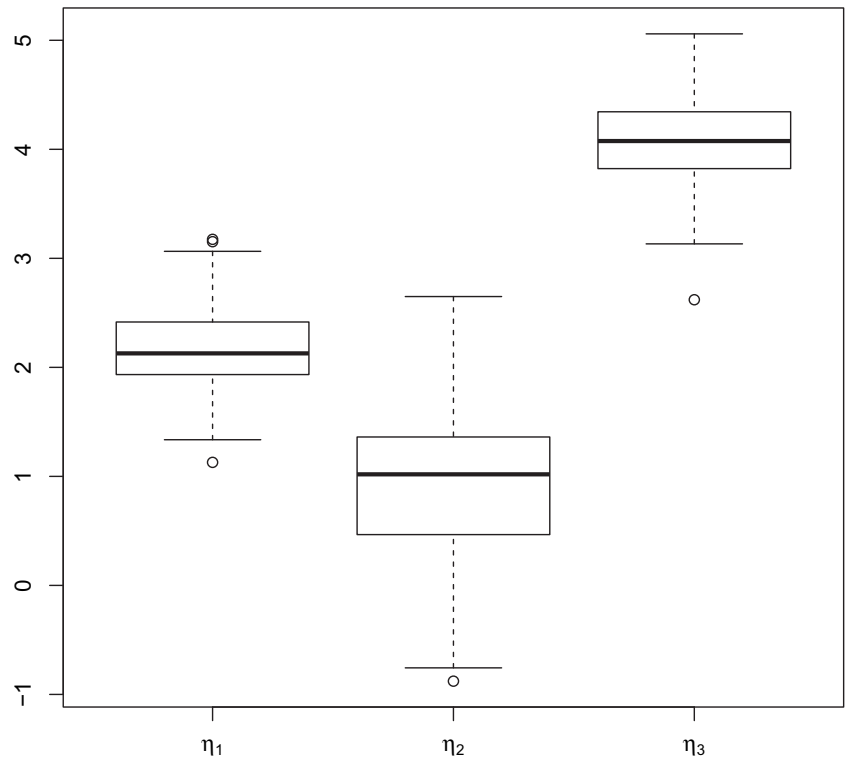

Fig. 10. Boxplots for each feature coefficient at the final time point during the WNS simulation experiment

6.2. Simulating management of white nose syndrome

The parameter values that were used in the generative model are given in Table 3. We set $\theta_3 = \theta_4$ and chose the magnitude of the treatment effect so that, if every location were treated at every time point during the management period, there would be a 95% reduction in the spread. However, when we simulate management of the disease, the strategies under consideration are limited to treating no more than 67 infected counties and 67 uninfected counties at each time point (which corresponds to treating at most 12% of the total locations). To evaluate the performance of the algorithm proposed, we simulate management of the disease from 2015 to 2022. As in Section 5, in addition to policy search, we also implement the proximal and myopic strategies. We use the same features for policy search as we did in Section 5. Also as in Section 5, in our implementation of Thompson sampling, we used draws from the estimated sampling distribution of the maximum likelihood estimator of θ to approximate draws from the posterior; at the first time point at which interventions are applied, the parameters $\theta_3 = \theta_4$ are not identified from the data so we sample them independently from an informative prior $N(4\theta_3, 1)$. Table 4 displays the average proportion of infected counties in the WNS simulation based on 100 Monte Carlo replications. The policy search algorithm resulted in significantly fewer infected counties than did competing methods. The myopic strategy is representable as in terms of the linear priority score using $\eta = (1, 0, 0)$; thus, one way to gain insight into the differences between the myopic policy and that of policy search is to examine the posterior distribution of η. Fig. 10 shows this posterior distribution. It can be seen that policy search puts large positive weights

on η_3, suggesting that it is putting a high priority on the secondary effects of infection relative to the myopic strategy.

The policy search algorithm, running on an Intel Xeon Server with 64 threads (3.4 GHz; 512 Gbytes DDR4 random-access memory), took an average of 20.4 min per Monte Carlo replication. This simulation demonstrates that policy search is a feasible and potentially powerful tool that can be used to inform the management of emerging infectious diseases like WNS.

7. Discussion

We proposed a statistical framework to study sequential treatment allocation in the context of managing emerging infectious diseases. On the basis of this framework, we identified several major computational and theoretical challenges that are associated with constructing an optimal treatment allocation strategy using accumulating data on disease spread. Among these challenges is interference between locations and subsequently exponential growth in the number of allocations as a function of locations. We used a low dimensional system dynamics model to impose (implicitly) structure on the nature of treatment interference and a prespecified class of strategies to reduce the computational complexity of searching for an optimal strategy. The proposed policy search estimator performed well in simulated experiments and shows promise as a means to inform management of emerging infectious diseases.

The framework proposed used some simplifying assumptions about the nature of the spatiotemporal treatment allocation problem that could be relaxed at the expense of additional modelling and/or computational complexity. We assumed that the state $\mathbf{S}^t$ was completely observed, without error at each location at each time point. In some applications, state observations may be sparse, noisy and irregularly spaced in time. Furthermore, the sampling design may be dependent on the evolution of the disease; for example, states might be sampled only at locations where an infection had been reported. Depending on the nature of the sampling design, it may be possible to combine sampling weights or imputation methods with policy search to estimate an optimal allocations strategy. Another important simplifying assumption is that of complete compliance of the decision maker, i.e. that the recommended allocations will actually be followed. Partial compliance, which is also known as partial controllability, could be incorporated in the framework proposed by adding a compliance model that described the distribution over allocations selected by the decision maker given the allocation recommended by policy search.

As indicated in our discussion of simplifying assumptions, we believe that the area of data-driven spatiotemporal treatment allocation is rife with important and exciting open problems. We briefly review several of the most pressing of these. Our proposed estimation algorithm relies on a postulated system dynamics model; an important extension is to construct semiparametric or non-parametric estimators of the optimal allocation strategy which are robust to misspecification of the system dynamics model. One potential approach to construct such estimators is to convert the Bellman equation for the optimal allocation strategy into an estimating equation (Maei *et al.*, 2010; Ertefaie, 2014). Another important direction for additional research is scaling spatiotemporal allocation algorithms to very large problems. Our current algorithm scales readily to settings with thousands of locations; however, additional computational considerations (both in terms of central processor unit clock cycles and memory) are needed to scale to larger problems. Finally, we believe that there is the potential for rich theoretical developments regarding the difference between cumulative expected utility under an estimated allocation strategy and the optimal strategy; this difference is known as the regret in the bandit algorithm literature (Robbins, 1952). Inspired by recent theoretical developments for Thompson

sampling (Korda *et al*., 2013), we believe that it will be possible to derive minimax-type regret bounds for algorithms similarly to our proposed policy search algorithm.

Acknowledgements

Any use of trade, product or firms' names is for descriptive purposes only and does not imply endorsement by the US Government.

References

Agrawal, S. and Goyal, N. (2011) Analysis of Thompson sampling for the multi-armed bandit problem. *Preprint arXiv:1111.1797*.

Agrawal, S. and Goyal, N. (2012) Analysis of Thompson sampling for the multi-armed bandit problem. In *Proc. Conf. Learning Theory*, vol. 23, pp. 39.1–39.26.

Agrawal, S. and Goyal, N. (2013) Thompson sampling for contextual bandits with linear payoffs. In *Proc. Int. Conf. Machine Learning*, vol. 3, pp. 127–135.

Anderson, R. M., May, R. M. and Anderson, B. (1992) *Infectious Diseases of Humans: Dynamics and Control*. Chichester: Wiley.

Banks, J. *et al*. (1998) *Handbook of Simulation*. New York: Wiley.

Barbu, C., Dumonteil, E. and Gourbière, S. (2009) Optimization of control strategies for non-domiciliated triatoma dimidiata, chagas disease vector in the Yucatán Peninsula, Mexico. *PLOS Negl. Trop. Dis.*, **3**, no. 4, article e416.

Barbu, C., Dumonteil, E. and Gourbière, S. (2011) Evaluation of spatially targeted strategies to control non-domiciliated triatoma dimidiata vector of Chagas disease. *PLOS Negl. Trop. Dis.*, **5**, no. 5, article e1045.

Bauch, C. T., Lloyd-Smith, J. O., Coffee, M. P. and Galvani, A. P. (2005) Dynamically modeling SARS and other newly emerging respiratory illnesses: past, present, and future. *Epidemiology*, **16**, 791–801.

Bernard, R. F., Foster, J. T., Willcox, E. V., Parise, K. L. and McCracken, G. F. (2015) Molecular detection of the causative agent of white-nose syndrome on Rafinesque's big-eared bats (corynorhinus rafinesquii) and two species of migratory bats in the southeastern USA. *J. Wldlif. Dis.*, **51**, 519–522.

Berry, D. A. and Fristedt, B. (1985) *Bandit Problems: Sequential Allocation of Experiments*. New York: Springer.

Bertsekas, D. P. (1996) *Neuro-dynamic programming*. Athena Scientific.

Bhatnagar, S., Prasad, H. and Prashanth, L. (2013) Kiefer-Wolfowitz algorithm. In *Stochastic Recursive Algorithms for Optimization*, pp. 31–39. New York: Springer.

Blatt, D., Murphy, S. A. and Zhu, J. (2004) A-learning for approximate planning. *Technical Report 04-63*, Methodology Center, Pennsylvania State University, State College.

Blehert, D. S., Hicks, A. C., Behr, M., Meteyer, C. U., Berlowski-Zier, B. M., Buckles, E. L., Coleman, J. T., Darling, S. R., Gargas, A. and Niver, R. (2009) Bat white-nose syndrome: an emerging fungal pathogen? *Science*, **323,** 227–227.

Borkar, V. S. (2008) *Stochastic Approximation*, vol. 1. New York: Cambridge University Press.

Bossenbroek, J. M., Kraft, C. E. and Nekola, J. C. (2001) Prediction of long-distance dispersal using gravity models: zebra mussel invasion of inland lakes. *Ecol. Appl.*, **11**, 1778–1788.

Boyles, J. G., Cryan, P. M., McCracken, G. F. and Kunz, T. H. (2011) Economic importance of bats in agriculture. *Science*, **332**, 41–42.

Bozzette, S. A., Boer, R., Bhatnagar, V., Brower, J. L., Keeler, E. B., Morton, S. C. and Stoto, M. A. (2003) A model for a smallpox-vaccination policy. *New Engl. J. Med.*, **348**, 416–425.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A. and Wiener, J. (2000) Graph structure in the web. *Comput. Netwrks*, **33**, 309–320.

Cesa-Bianchi, N. and Lugosi, G. (2006) *Prediction, Learning, and Games*. Cambridge: Cambridge University Press.

Chakraborty, B. and Moodie, E. E. (2013) *Statistical Methods for Dynamic Treatment Regimes*. New York: Springer.

Chapelle, O. and Li, L. (2011) An empirical evaluation of Thompson sampling. In *Advances in Neural Information Processing Systems*, pp. 2249–2257.

Cornelison, C. T., Keel, M. K., Gabriel, K. T., Barlament, C. K., Tucker, T. A.., Pierce, G. E. and Crow, S. A. (2014) A preliminary report on the contact-independent antagonism of pseudogymnoascus destructans by rhodococcus rhodochrous strain dap96253. *BMC Micrbiol.*, **14**, no. 1, article 246.

Diez Roux, A. V. (2004) Estimating neighborhood health effects: the challenges of causal inference in a complex world. *Socl Sci. Med.*, **58**, 1953–1960.

Drake, J. M. and Lodge, D. M. (2004) Global hot spots of biological invasions: evaluating options for ballast–water management. *Proc. R. Soc. Lond.* B, **271**, 575–580.

Ertefaie, A. (2014) Constructing dynamic treatment regimes in infinite-horizon settings. *Preprint arXiv:1406.0764*.
Estrada, E. and Rodriguez-Velazquez, J. A. (2005) Subgraph centrality in complex networks. *Phys. Rev.* E, **71**, no. 5, article 056103.
Ferguson, N. M., Donnelly, C. A. and Anderson, R. M. (2001a) The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science*, **292**, 1155–1160.
Ferguson, N. M., Donnelly, C. A., and Anderson, R. M. (2001b) Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature*, **413**, 542–548.
Ferguson, N. M., Keeling, M. J., Edmunds, W. J., Gani, R., Grenfell, B. T., Anderson, R. M. and Leach, S. (2003) Planning for smallpox outbreaks. *Nature*, **425**, 681–685.
Field, K., Reeder, S., Rogers, E., James, M., Sigler, L., Vodzak, M. Moore, M., Johnson, J. and Reeder, D. (2014) Anti-fungal immune responses to pseudogymnoasces destructans in bats affected by white-nose syndrome (vet2p. 1041). *J. Immun.*, **192**, suppl. 1, 207–213.
Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2014) *Bayesian Data Analysis*, vol. 2. New York: Taylor and Francis.
Ghavamzadeh, M., Mannor, S., Pineau, J., Tamar, A. *et al*. (2015) *Bayesian Reinforcement Learning: a Survey*. Singapore: World Scientific.
Ghosal, S. and van der Vaart, A. (2017) *Fundamentals of Nonparametric Bayesian Inference*. New York: Cambridge University Press.
Goldberg, Y. and Kosorok, M. R. (2012) Q-learning with censored data. *Ann. Statist.*, **40**, 529.
Gopalan, A. and Mannor, S. (2015) Thompson sampling for learning parameterized Markov decision processes. In *Proc. 28th Conf. Learning Theory*, pp. 861–898.
Gopalan, A., Mannor, S. and Mansour, Y. (2014) Thompson sampling for complex online problems. In *Proc. Int. Conf. Machine Learning*, vol. **14**, pp. 100–108.
Gosavi, A. (2003) *Simulation-based Optimization: Parametric Optimization Techniques and Reinforcement Learning*. New York: Springer.
Halloran, M. E. and Struchiner, C. J. (1995) Causal inference in infectious diseases. *Epidemiology*, **6**, 142–151.
Hollingsworth, T. D. (2009) Controlling infectious disease outbreaks: lessons from mathematical modelling. *J. Publ. Hlth Poly*, **30**, 328–341.
Hollingsworth, T. D., Ferguson, N. M. and Anderson, R. M. (2006) Will travel restrictions control the international spread of pandemic influenza? *Nat. Med.*, **12**, 497–499.
Hong, G. and Raudenbush, S. W. (2006) Evaluating kindergarten retention policy. *J. Am. Statist. Ass.*, **101**, 901–910.
Hoyt, J. R., Cheng, T. L., Langwig, K. E., Hee, M. M., Frick, W. F. and Kilpatrick, A. M. (2015) Bacteria isolated from bats inhibit the growth of pseudogymnoascus destructans, the causative agent of white-nose syndrome. *PLOS One*, **10**, no. 4, article e0121329.
Huang, C.-Y., Sun, C.-T., Hsieh, J.-L. and Lin, H. (2004) Simulating SARS: small-world epidemiological modeling and public health policy assessments. *J. Artif. Soc. Socl Simuln*, **7**, no. 4.
Hudgens, M. G. and Halloran, M. E. (2008) Toward causal inference with interference. *J. Am. Statist. Ass.*, **103**, 832–842.
Hufnagel, L., Brockmann, D. and Geisel, T. (2004) Forecast and control of epidemics in a globalized world. *Proc. Natnl. Acad. Sci. USA*, **101**, 15124–15129.
Jacquez, J. A., Simon, C. P., Koopman, J., Sattenspiel, L. and Perry, T. (1988) Modeling and analyzing HIV transmission: the effect of contact patterns. *Math. Biosci.*, **92**, 119–199.
Jung, E., Iwami, S., Takeuchi, Y. and Jo, T.-C. (2009) Optimal control strategy for prevention of avian influenza pandemic. *J. Theoret. Biol.*, **260**, 220–229.
Kaelbling, L. P., Littman, M. L. and Moore, A. W. (1996) Reinforcement learning: a survey. *J. Artif. Intell. Res.*, **4**, 237–285.
Kang, C., Janes, H. and Huang, Y. (2014) Combining biomarkers to optimize patient treatment recommendations. *Biometrics*, **70**, 695–707.
Kaplan, E. H., Craft, D. L. and Wein, L. M. (2002) Emergency response to a smallpox attack: the case for mass vaccination. *Proc. Natnl. Acad. Sci. USA*, **99**, 10935–10940.
Kaufmann, E., Korda, N. and Munos, R. (2012) Thompson sampling: an asymptotically optimal finite-time analysis. In *Algorithmic Learning Theory*, pp. 199–213. New York: Springer.
Keele, L. (2015) The statistics of causal inference: a view from political methodology. *Polit. Anal.*, **23**, 313–335.
Keeling, M. J. (2005) Models of foot-and-mouth disease. *Proc. R. Soc. Lond.* B, **272**, 1195–1202.
Keeling, M. J. and Rohani, P. (2011) *Modeling Infectious Diseases in Humans and Animals*. Princeton: Princeton University Press.
Kilpatrick, A. M. (2011) Globalization, land use, and the invasion of West Nile virus. *Science*, **334**, 323–327.
Kohn, K. W. (1999) Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Molec. Biol. Cell*, **10**, 2703–2734.
Kolaczyk, E. D. (2009) *Statistical Analysis of Network data*. New York: Springer.
Korda, N., Kaufmann, E. and Munos, R. (2013) Thompson sampling for 1-dimensional exponential family bandits. In *Advances in Neural Information Processing Systems*, pp. 1448–1456.

Korenromp, E. L., Van Vliet, C., Grosskurth, H., Gavyole, A., Van der Ploeg, C. P., Fransen, L., Hayes, R. J. and Habbema, J. D. F. (2000) Model-based evaluation of single-round mass treatment of sexually transmitted diseases for HIV control in a rural African population. *Aids*, **14**, 573–593.
Kramer, A. M., Pulliam, J. T., Alexander, L. W., Park, A. W., Rohani, P.. and Drake, J. M. (2016) Spatial spread of the West Africa Ebola epidemic. *Open Sci.*, **3**, no. 8, article 160294.
Kretzschmar, M., Van den Hof, S., Wallinga, J. and Van Wijngaarden, J. (2004) Ring vaccination and smallpox control. *Emergng Infect. Dis.*, **10**, no. 5, 832.
Kushner, H. J. and Yin, G. (2003) *Stochastic Approximation and Recursive Algorithms and Applications*. New York: Springer.
Laber, E. B., Linn, K. A. and Stefanski, L. A. (2014) Interactive model building for q-learning. *Biometrika*, **101**, 831–847.
Laber, E. and Zhao, Y. (2015) Tree-based methods for estimating individualized treatment regimes. *Biometrika*, **102**, 501–514.
Law, A. M., Kelton, W. D. and Kelton, W. D. (1991) *Simulation Modeling and Analysis*, vol. 2. New York: McGraw-Hill.
Lekone, P. E. and Finkenstädt, B. F. (2006) Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics*, **62**, 1170–1177.
Le Menach, A., Vergu, E., Grais, R. F., Smith, D. L. and Flahault, A. (2006) Key strategies for reducing spread of avian influenza among commercial poultry holdings: lessons for transmission to humans. *Proc. R. Soc. Lond.* B, **273**, 2467–2475.
Li, S.-L., Bjørnstad, O. N., Ferrari, M. J., Mummah, R., Runge, M. C., Fonnesbeck, C. J., Tildesley, M. J., Probert, W. J. and Shea, K. (2017) Essential information: uncertainty and optimal control of Ebola outbreaks. *Proc. Natnl. Acad. Sci. USA*, **114**, 5659–5664.
Liu, R. Y. (1990) On a notion of data depth based on random simplices. *Ann. Statist.*, **18**, 405–414.
Ma, Z., Zhou, Y. and Wu, J. (2009) *Modeling and Dynamics of Infectious Diseases*, vol. 11. Singapore: World Scientific.
Maei, H. R., Szepesvári, C., Bhatnagar, S. and Sutton, R. S. (2010) Toward off-policy learning control with function approximation. In *Proc. 27th Int. Conf. Machine Learning*, pp. 719–726.
Mahadevan, S. (2009) *Learning Representation and Control in Markov Decision Processes*. Now.
Maher, S. P., Kramer, A. M., Pulliam, J. T., Zokan, M. A., Bowden, S. E., Barton, H. D., Magori, K. and Drake, J. M. (2012) Spread of white-nose syndrome on a network regulated by geography and climate. *Nat. Communs*, **3**, 1306.
Murphy, S. A. (2003a) Optimal dynamic treatment regimes. *J. R. Statist. Soc.* B, **65**, 331–355.
Murphy, S. A. (2003b) Optimal dynamic treatment regimes (with discussion). *J. R. Statist. Soc.* B, **65**, 331–366.
Murphy, S. A. (2005) A generalization error for Q-learning. *J. Mach. Learn. Res.*, **6**, 1073–1097.
Newey, W. K. (1997) Convergence rates and asymptotic normality for series estimators. *J. Econmetr.*, **79**, 147–168.
O'Donoghue, A. J., Knudsen, G. M., Beekman, C., Perry, J. A., Johnson, A. D., DeRisi, J. L., Craik, C. S. and Bennett, R. J. (2015) Destructin-1 is a collagen-degrading endopeptidase secreted by pseudogymnoascus destructans, the causative agent of white-nose syndrome. *Proc. Natnl. Acad. Sci. USA*, **112**, 7478–7483.
Ogburn, E. L. and VanderWeele, T. J. (2014) Vaccines, contagion, and social networks. *Preprint arXiv:1403.1241*.
Orellana, L., Rotnitzky, A. and Robins, J. (2010) Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part i: main content. *Int. J. Biostatist.*, **6**, no. 2, 1–49.
Osband, I., Russo, D. and Van Roy, B. (2013) (More) efficient reinforcement learning via posterior sampling. In *Advances in Neural Information Processing Systems*, pp. 3003–3011.
Pearl, J. (2010) On the consistency rule in causal inference: axiom, definition, assumption, or theorem? *Epidemiology,* **21,** 872–875.
Powell, W. B. (2007) *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. New York: Wiley.
Rich, B. (2013) Optimal dynamic treatment regime structural nested mean models: improving efficiency through diagnostics and re-weighting and application to adaptive individualized dosing. *Dissertation*. McGill University, Montreal.
Rich, B., Moodie, E. E. M., Stephens, D. A. and Platt, R. W. (2014) Simulating sequential multiple assignment randomized trials to generate optimal personalized warfarin dosing strategies. *Clin. Trials*, **11**, 435–444.
Riley, S. (2007) Large-scale spatial-transmission models of infectious disease. *Science*, **316**, 1298–1301.
Robbins, H. (1952) Some aspects of the sequential design of experiments. *Bull. Am. Math. Soc.* **58**, 527–535.
Robins, J. M. (2004a) Optimal structural nested models for optimal sequential decisions. In *Proc. 2nd Seattle Symp. Biostatistics*, pp. 189–326. New York: Springer.
Robins, J. (2004b) Optimal structural nested models for optimal sequential decisions. In *Proc. 2nd Seattle Symp. Biostatistics: Analysis of Correlated Data*. New York: Springer.
Robins, J., Orellana, L. and Rotnitzky, A. (2008) Estimation and extrapolation of optimal treatment and testing strategies. *Statist. Med.*, **27**, 4678–4721.
Rubin, D. (1978) Bayesian inference for causal effects: the role of randomization. *Ann. Statist.*, **6**, 34–58.
Russo, D. and Van Roy, B. (2014) An information-theoretic analysis of Thompson sampling. *J. Mach. Learn. Res.*, **17**, 1–30.

Schulte, P., Tsiatis, A., Laber, E. and Davidian, M. (2014) Q- and a-learning methods for estimating optimal dynamic treatment regimes. *Statist. Sci.*, **29**, 640–661.
Scott, S. L. (2010) A modern bayesian look at the multi-armed bandit. *Appl. Stochast. Modls Bus. Indstry*, **26**, 639–658.
Sen, A. and Smith, T. (2012) *Gravity Models of Spatial Interaction Behavior*. New York: Springer Science and Business Media.
Sobel, M. E. (2006) What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference. *J. Am. Statist. Ass.*, **101**, 1398–1407.
Spall, J. C. (2005) *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. New York: Wiley.
Splawa-Neyman, J., Dabrowska, D., Speed, T. *et al*. (1990) On the application of probability theory to agricultural experiments: essay on principles, section 9. *Statist. Sci.*, **5**, 465–472.
Strogatz, S. H. (2001) Exploring complex networks. *Nature*, **410**, 268–276.
Subcommittee on Fisheries, Wildlife, and Oceans (2011) *Why We Should Care about Bats: Devastating Impact White-nose Syndrome is having on One of Nature's Best Pest Controllers*. Committee on Natural Resources.
Sugiyama, M. (2015) *Statistical Reinforcement Learning: Modern Machine Learning Approaches*. Boca Raton: CRC Press.
Sutton, R. and Barto, A. (1998) *Reinforcment Learning: an Introduction*. Cambridge: Massachusetts Institute of Technology Press.
Szymanski, J. A., Runge, M. C., Parkin, M. J. and Armstrong, M. (2009) White-nose syndrome management: report on structured decision making initiative. *Report*. Department of the Interior, US Fish and Wildlife Service, Fort Snelling.
Thompson, W. R. (1933) On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, **25**, 285–294.
Tildesley, M. J., Savill, N. J., Shaw, D. J., Deardon, R., Brooks, S. P., Woolhouse, M. E., Grenfell, B. T. and Keeling, M. J. (2006) Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. *Nature*, **440**, 83–86.
Truscott, J. and Ferguson, N. M. (2012) Evaluating the adequacy of gravity models as a description of human mobility for epidemic modelling. *PLOS Computnl Biol.*, **8**, no. 10, article e1002699.
Turner, J. M., Warnecke, L., Wilcox, A., Baloun, D., Bollinger, T. K., Misra, V. and Willis, C. K. (2015) Conspecific disturbance contributes to altered hibernation patterns in bats with white-nose syndrome. *Physiol. Behav.*, **140**, 71–78.
US Fish and Wildlife Service (2015) White-nose syndrome: a coordinated response to the devastating bat disease. US Fish and Wildlife Service, Fort Snelling.
VanderWeele, T. J., Hernan, M. A. *et al*. (2013) Causal inference under multiple versions of treatment. *J Causl Inf.*, **1**, 1–20.
VanderWeele, T. J. and Tchetgen Tchetgen, E. J. (2011) Effect partitioning under interference in two-stage randomized vaccine trials. *Statist. Probab. Lett.*, **81**, 861–869.
Williams, R. J. and Martinez, N. D. (2000) Simple rules yield complex food webs. *Nature*, **404**, 180–183.
Xia, Y., Bjørnstad, O. N. and Grenfell, B. T. (2004) Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *Am. Natlst*, **164**, 267–281.
Xu, X., Kypraios, T. and O'Neill, P. D. (2016a) Bayesian non-parametric inference for stochastic epidemic models using Gaussian processes. *Biostatistics*, **17**, 619–633.
Xu, Y., Müller, P., Wahed, A. S. and Thall, P. F. (2016b) Bayesian nonparametric estimation for dynamic treatment regimes with sequential transition times. *J. Am. Statist. Ass.*, **111**, 921–950.
Zhang, B., Tsiatis, A. A., Davidian, M., Zhang, M. and Laber, E. (2012a) Estimating optimal treatment regimes from a classification perspective. *Stat*, **1**, 103–114.
Zhang, B., Tsiatis, A. A., Laber, E. B. and Davidian, M. (2012b) A robust method for estimating optimal treatment regimes. *Biometrics*, **68**, 1010–1018.
Zhang, B., Tsiatis, A. A., Laber, E. B. and Davidian, M. (2013) Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika*, **100**, 681–694.
Zhao, Y., Zeng, D., Laber, E. B. and Kosorok, M. R. (2014a) Doubly robust learning for estimating individualized treatment with censored data. *Biometrika*, to be published.
Zhao, Y., Zeng, D., Laber, E. B. and Kosorok, M. R. (2014b) New statistical learning methods for estimating optimal dynamic treatment regimes. *J. Am. Statist. Ass.*, to be published.
Zhao, Y., Zeng, D., Rush, A. J. and Kosorok, M. R. (2012) Estimating individualized treatment rules using outcome weighted learning. *J. Am. Statist. Ass.*, **107**, 1106–1118.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Supplementary material'.