University of California
Santa Barbara

Epidemic Detection in Two Populations

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Statistics and Applied Probability

by

Katherine Shatskikh

Committee in charge:

Professor Mike Ludkovski, Chair
Professor Raya Feldman
Professor John Hsu

March 2017

The Dissertation of Katherine Shatskikh is approved.

Professor Raya Feldman

Professor John Hsu

Professor Mike Ludkovski, Committee Chair

March 2017

Epidemic Detection in Two Populations

To My Grandmother

Acknowledgements

First and foremost, I would like to thank my advisor Professor Michael Ludkovski, who has provided me with invaluable professional and academic advice. I am forever grateful for your continuous support and patience over the years.

I would also like to thank Professor John Hsu, Professor Raya Feldman, Professor Wendy Meiring, and Professor Drew Carter for sharing their knowledge during classes as well as mentoring throughout my teaching career in UCSB.

Last, but not least, I would like to thank Michael Nava and Mark Dela for interesting conversations and emotional support.

This material is based upon work partially supported by the National Science Foundation under Grant No. ATD - 1222262.

Curriculum Vitæ

Katherine Shatskikh

Education

2017	Ph.D. in Statistics and Applied Probability (Expected), University of California, Santa Barbara.
2012	M.A. in Mathematical Statistics, University of California, Santa Barbara.
2009	B.S. in Economics, Magnitogorsk State Technical University, Magnitogorsk, Russia

Publications

Computational Method for Epidemic Detection in Multiple Populations, with M. Ludkovski, Online Journal of Public Health Informatics. 2015

Bayesian Epidemic Detection in Multiple Populations, with M. Ludkovski, 2015, Preprint.

Abstract

Epidemic Detection in Two Populations

by

Katherine Shatskikh

Traditional epidemic detection algorithms make decisions using only local information. We propose a novel approach that explicitly models spatial information fusion from several meta-populations. Our method also takes into account cost-benefit considerations regarding the announcement of epidemic. We utilize a compartmental stochastic model within a Bayesian detection framework, which leads to a dynamic optimization problem. The resulting adaptive, non-parametric detection strategy optimally balances detection delay vis-a-vis probability of false alarms. Our algorithm can also be used to optimize an existing detection strategy. Taking advantage of the underlying state-space structure, we represent the stopping rule in terms of a detection map, which visualizes the relationship between the multivariate system state and policy making. It also allows us to obtain an efficient simulation-based solution algorithm that is based on the Sequential Regression Monte Carlo (SRMC) approach of Gramacy and Ludkovski (SIFIN, 2015). We present two models for pseudo-posterior and illustrate our results on two synthetic examples. We also quantify the advantages of our adaptive detection relative to conventional threshold-based strategies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

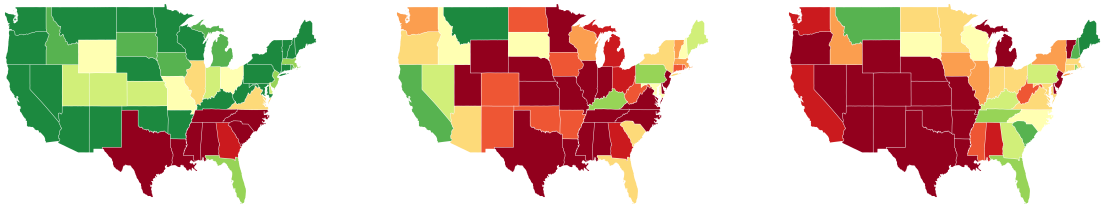
Infectious disease epidemics intrinsically unfold across both space and time. As a result, biosurveillance algorithms need to integrate spatio-temporal data. This is especially so in the context of statistical inference, whereby syndromic surveillance at neighboring locales carries additional information that can be fused for improved decision making in terms of initiating and organizing epidemic counter-measures. A crucial first step for response strategies is to identify, or detect, in real-time the epidemic outset. In this thesis, we propose a methodology that allows for such *optimal* decision-making with spatial information fusion. Specifically, we investigate a model that combines quickest detection with a spatial metapopulation setup, integrating information received from multiple geographic domains. To reflect the inherent uncertainty in epidemic evolution (which is amplified under partial information), we develop a stochastic compartmental (or state-space) epidemic model, which allows us to generate adaptive, nonparametric detection rules. Extant approaches largely propose heuristic detection strategies, concentrating primarily on the inferential aspect of the statistical model [Chan et al., 2010, Lin and Ludkovski, 2014, Sheinson et al., 2014, Skvortsov and Ristic, 2012]. For instance, a typical approach is to announce an epidemic as soon as the estimated number of infecteds in

the local population is above a fixed $\bar{I}$. In contrast, we dynamically optimize the detection strategy, to come up with a "best" detection rule within our mechanistic outbreak model.

Traditional compartmental epidemic models deal with a single population; the spatial aspect is treated by building a series of such single-population models that are estimated/forecasted independently. This is also a common surveillance approach, especially for recurring infectious epidemics, such as influenza-like illness (ILI), dengue fever, or measles. For example, in the US the existing biosurveillance systems for flu operate primarily at the state level and are siloed across states. This limitation of existing practice was brought into sharp relief during the 2014 Ebola outbreak in West Africa. The epidemic has been accompanied by a dearth of reliable information, leading to extreme spread in forecasts regarding the future course of the outbreak. In addition, numerous statistical methods [WHO, 2014, Chowell and Nishiura, 2014, Fisman et al., 2014] were put forth attempting to infer in "real-time" the actual size and parameters of the outbreak in different locales. However, nearly all these methods were single-population, so that when trying for example to infer the number of Ebola infecteds in Liberia, only Liberian data was utilized, completely ignoring similar and highly relevant data from neighboring Guinea and Sierra Leone. Similarly, at the more granular provincial level, data from neighboring provinces was generally not used during estimation procedures.

For a less dramatic and perhaps more statistically convenient example, we discuss the yearly influenza outbreaks in United States. Figure 1.1 illustrates the spatial dynamics of ILI during the 2012-13 flu season. Observe that the peak of the outbreak varied significantly (up to 6-8 weeks difference) across different parts of the country. Nevertheless, there is a clear propagation of the outbreak, making spatial information fusion desirable. Figure 1.1 indicates that the current, single-population based detection protocols are not sufficient; for instance the fact that there are increased ILI levels in Arizona is ought to

be taken into account when trying to detect or forecast the epidemic start in California. A further important remark is that the illustrated spatial spread is year-specific, and in other years rather different patterns may be observed.



Week 49, 2012 Week 1, 2013 Week 4, 2013

Figure 1.1: Spread of Influenza during the 2012-13 Flu season according to FluView CDC data. The colors represent weekly ILI activity levels in terms of percentage of doctor visits attributed to ILI relative to low-season baseline. Green indicates at/below mean, while shades of red indicate outbreak activity (with darkest color corresponding to eight or more standard deviations above the mean). Weeks are numbered from January 1, and are 12/3-9/2012 (Week 49), 12/31/2012-1/6/2013 (Week 1) and 1/21-27/2013 (Week 4), respectively. Data source: `http://www.cdc.gov/flu/weekly/pastreports.htm`.

1.1 Contributions

In this thesis we formulate and analyze an epidemic detection problem within a multi-population paradigm. To do so, we develop a reduced compartmental model that extends the classical Susceptible-Infected-Recovered (SIR) setup to two population pools. Pools are interpreted as distinct geographic regions, e.g. states or counties. To fix ideas, we consider the situation where the epidemic begins in Pool 1 and subsequently may be transmitted to Pool 2 via infecteds that travel between the two pools. The aim of the policy-maker is to detect, as soon as possible and in online fashion, the onset of epidemic in Pool 2.

To capture the inferential aspect, we assume that full information is available about the outbreak in Pool 1, but only partial information about Pool 2. This situation can

happen if Pool 1 has a better surveillance system than Pool 2. As a result, one has to make imperfect decisions and in particular address the canonical trade-off between making announcements too early (so called "false alarms") and making decisions too late ("detection delay"). Indeed, if the detection is too late, then a certain number of infections would be missed and it would be harder to stop the epidemic from spreading. If the detection is premature, human, financial and reputational resources would be wasted. Therefore, a careful trade-off between those costs should be done to balance costs due to epidemic morbidity and costs arising from policy actions. We then use the above cost analysis to quantify decision-making quality and to define optimality of detection strategies.

Mathematically, we cast the online detection problem as a dynamic optimization problem, connecting to the classical dynamic programming formulation [Bertsekas, 2005] in control theory. A major challenge with dynamic programming (which is perhaps the prime reason for the lack in its uptake in the biosurveillance community) is computational bottlenecks due to the curse of dimensionality. Indeed, the above optimization problem is nontrivial from several directions. First, because the underlying system is stochastic, the optimal solution is *adaptive*, i.e. a function of the current system state. Consequently, there is no simple description to the resulting detection strategy which is instead summarized through a detection *map* that translates system states into optimal detection decisions. Second, the nonlinear dynamics of the SIR model preclude analytic solutions. Crucially, there are no analytic expressions for the future distribution of the system state, which necessitates the use of numerical approximations to solve the optimization problem. Third, because the system state is multivariate and too large to enumerate, the corresponding integrals are computationally demanding.

However, taking advantage of the detection strategy structure, which requires simply announcing at each stage whether the epidemic has reached Pool 2 or not, we implement

an efficient numerical algorithm. Specifically, we rely on the recent Sequential Regression Monte Carlo (SRMC) method of Gramacy and Ludkovski [2015], which blends modern statistical tools, including nonparametric regression and sequential design, with approximate dynamic programming, to drastically mitigate issues of computational efficiency.

1.2 Spatial Stochastic Epidemic Models

Mathematical models of infectious disease epidemics have become an important tool in the arsenal of public health policy. In an idealized world, detection reduces to the mathematical problem of clustering, tracking the health status of the surveyed individuals and identifying unusual aberrations in either the temporal or spatial dimensions. In reality, there is the additional aspect of missing information which necessitates the application of statistical inference algorithms, as well as a mathematical model for the epidemic. In the context of online inference, a simple mechanistic approach that allows for maximum tractability continues to be the most popular, and is also adopted here. Specifically, we rely on the formalism of an SIR model [Andersson and Britton, 2000] that implies proportional homogenous mixing between infecteds and susceptibles within a population pool. Spatial heterogeneity is captured by incorporating meta-populations, also known as patch models [Ball and Clancy, 1993, Allen et al., 2009, Neal, 2012]. The multi-patch approach partitions the global population into distinct discrete regions or pools, allowing for local spread of the epidemic within each pool, as well as global transmission that is specified via a mobility matrix. As in Ball and Clancy [1993], Neal [2012] we assume that susceptibles are stationary, while infecteds can move or travel between the pools, creating cross-infections.

Alternative frameworks for epidemic spread include point process models [Neill, 2011], and network models [Keeling and Eames, 2005] that provide more nuanced interaction

between individuals to mimic existing social structures, such as households, schools, and workplaces. At even more detail, agent-based models [Ajelli et al., 2010] generate micro-simulations that provide a detailed synthetic view for each individual and their social interactions. Such models can also incorporate precise travel patterns [Rvachev and Longini, 1985]. However, the latter paradigms are geared towards realistic *forecasting* of epidemic progress and are less suited for online detection due to intractable inference in terms of observed data and the computational expenses in generating micro-scenarios.

A variety of approaches exist for constructing outbreak detection rules, see for example the recent survey by Shmueli and Burkom [2010], and the monograph by Lawson [2013]. Quality control methods [Cowling et al., 2006] introduced in the 1950s form the simplest class of rules and continue to be common. Other heuristics include moving-average tools [Zhou and Lawson, 2008], various scan statistics [Kulldorff et al., 2005, Neill, 2011], and branching-process approximations [Nishiura, 2011]. More explicit cost-benefit analysis for the trade-off between false alarms and detection delay can be applied using the Cumulative Sum (CUSUM) framework [Yang et al., 2017]. CUSUM also underlies the early aberration response system (EARS) employed by the Centers for Disease Control [Hutwagner et al., 2005]. Alternatively, Bayesian methods allow to further assess the uncertainty involved in decision-making based on partial information. Two main types are hidden Markov models [LeStrat and Carrat, 1999, Martínez-Beneito et al., 2008] and Bayesian hierarchical models [Chan et al., 2010, Sebastiani et al., 2006]. The Bayesian paradigm translates epidemic data into the posterior probability of an outbreak. To convert the latter into a detection rule, one typically employs a simple threshold strategy. For example, in Chan et al. [2010], the authors recommend "an alert for action if the posterior probability is larger than 70%". We further refine this approach by deriving *optimal*, non-parametric detection strategies based on the inputted cost-benefit parameters.

Detection can be seen as a basic form of epidemic response, and indeed our computational methodology can be extended to this more general problem. In that sense, this thesis extends Ludkovski's previous work on stochastic control methods for controlling epidemics [Ludkovski and Niemi, 2011, 2010]. Similar to Ludkovski and Niemi [2011], we design a Bayesian dynamic optimization algorithm for biosurveillance decision policy. Other mathematically oriented studies that consider optimal control of epidemics include Tanner et al. [2008], Wearing et al. [2005].

In the context of detection with limited information, a spatial epidemic model requires information fusion. Fusion of information channels for the purpose of biosurveillance has been an area of intense research in the past decade. On the one hand, novel information sources, such as social media [Culotta, 2010] or internet data [Dukic et al., 2012] have created new opportunities for syndromic surveillance. On the other hand, developments in statistical fusion techniques [Shmueli and Burkom, 2010, Banks et al., 2012, Noufaily et al., 2013] have led to new ways of integrating multivariate information streams. In particular, there has been a lot of interest in online Bayesian approaches [Lin and Ludkovski, 2014, Sheinson et al., 2014, Nishiura, 2011, Dukic et al., 2012, Yang et al., 2014] that allow for predictive modeling and forecasting of epidemics. The above models all focus on a single homogenous population with the different surveillance channels complementing each other. In contrast, we consider multiple underlying population pools each with a distinct, but co-dependent information channel. In terms of explicitly accounting for spatial propagation, our work is closest to Ludkovski [2012] who considered a spatial "wave" model for an epidemic. In the present thesis, we connect this framework to the SIR context, modeling epidemic spread across geographically-based population pools. The resulting decision strategy provides insights into integrating data from multiple spatial locales for the purposes of detection, cf. Chapter 7 below.

In Chapter 2 we set up a detection problem via Quickest Detection method. In

Chapter 3 we will talk about stochastic epidemic models. However since only partial information available about the epidemic in Pool 2, we will construct two reduced models in Chapter 4. Chapter 5 discusses the computational methods of solving our detection problem defined in Chapter 2. In Chapter 6 we explore two case studies based on the two reduced models we constructed. In Chapter 7 we summarize our findings and propose potential improvements to our algorithm.

Remark 1 *The parts of this thesis, which are related to the First Reduced Model defined in Chapter 4 (Chapter 2, Sections 3.1.1 and 3.2 of Chapter 3, Chapter 4 without Section 4.2, Chapter 5, and Section 6.1), has already appeared in the paper of Shatskikh and Ludkovski [2015] and in the conference proceedings Shatskikh and Ludkovski [2015]. In this thesis we add the construction of the Second Reduced Model (defined in Section 4.2) as well as the Case Study (Section 6.2) based on the UK data (Section 3.3). The material based on Appendix A is getting ready to appear in the paper of Shatskikh and Ludkovski [2017].*

Chapter 2

Quickest Detection

A state-space model $\mathfrak{X}_t$ describes the epidemic state at times $t = 0, 1, 2, \ldots$. A typical length of one time period in biosurveillance is a week. The precise components of $\mathfrak{X}$ will be specified in Chapter 4; abstractly $\mathfrak{X}$ is taken to be a stochastic Markov process taking values in a state space $\mathcal{X} \subset \mathbb{R}^d$ and summarizes information about both Pool 1 and Pool 2. In particular, $\mathfrak{X}$ contains information about the number of infecteds $I_t^{(k)}$ in Pool $k = 1, 2$ at time t. The transition kernel of $\mathfrak{X}$ is assumed to be time-stationary and is denoted by $p_s(\mathbf{x}|\mathbf{y}) \equiv P(\mathfrak{X}_{t+s} = \mathbf{x}|\mathfrak{X}_t = \mathbf{y})$, $\mathbf{x}, \mathbf{y} \in \mathcal{X}$.

The aim of the policy maker is to detect the onset of epidemic in Pool 2. A detection strategy is probabilistically represented as a dynamic "alarm" which announces an outbreak in Pool 2 based on information gathered so far. Only a single announcement is allowed; once announced, the detection problem is assumed to be over. The set of such detection strategies is expressed through the set $\mathcal{S}$ of $\mathcal{F}$-stopping times, where $\mathcal{F}_t = \sigma(\mathfrak{X}_{0:t})$ is the information filtration generated by $\mathfrak{X}$ by time t. A strategy $\tau \in \mathcal{S}$ is a random variable taking values in $\tau \in \{0, 1, 2, \ldots\}$, such that $\{\tau = t\} \in \mathcal{F}_t$ (this requirement captures the fact that τ must be "online" in terms of the information available so far). Thanks to the Markov property of $\mathfrak{X}$, the structure of τ can be summarized

via a detection map. Indeed, at each time-step there is the binary decision to either "announce" an outbreak (subset $\mathfrak{S}$), or wait for another period (subset $\mathfrak{C}$). Since the evolution of $\mathfrak{X}$ is stationary in time, the corresponding partition of the state space is also independent of t. Dynamically, this implies that τ announces the epidemic the first time that the state $\mathfrak{X}$ enters the region $\mathfrak{S} \subset \mathcal{X}$,

$$\tau = \inf\{t : \mathfrak{X}_t \in \mathfrak{S}\}. \tag{2.1}$$

Equation (2.1) gives a one-to-one correspondence between detection strategies τ and detection maps $\mathfrak{S}$. In other words, the detection strategies we consider are of online feedback type, based on the trajectory of $\mathfrak{X}$.

As mentioned, the dynamic optimization objective consists in optimally trading off the concern of premature announcements against any potential delays. These conflicting costs are measured through the immediate stopping cost $d(\mathbf{x})$ and the cost of waiting. The immediate costs are linked to the penalty for false alarms, specified by a given constant C_{FA}. We assume that C_{FA} is paid if and only if the epidemic has not yet reached Pool 2, so that

$$d(\mathbf{x}) := C_{\mathrm{FA}} \cdot \mathbf{1}_{\{I_0^{(2)} \leq \bar{I}\}}, \tag{2.2}$$

where $\bar{I}$ is a fixed threshold for which it is assumed that there is no epidemic and $I_0^{(2)}$ is the number of infected individuals at time 0 . The value of $\bar{I}$ is usually computed from the historical data [Farrington et al., 1996, Guintran et al., 2006]. Waiting costs are assumed to be proportional to detection delay, i.e. the time between the outbreak reaching Pool 2 and outbreak announcement. Define θ to be the time when the second

population gets infected from the first population, i.e.

$$\theta := \inf\{t : I_t^{(2)} > \bar{I}\},$$

where $\bar{I}$ is defined as in Equation 2.2 and $I_t^{(2)}$ is the number of infected individuals at time t. Then the detection delay is $\max(\tau - \theta, 0)$ and carries cost $C_{\text{Delay}} \max(\tau - \theta, 0)$. This is equivalent to charging waiting costs of $C_{\text{Delay}} \mathbf{1}_{\{I_t^{(2)} > \bar{I}\}}$ at each step until surveillance is terminated at the random instant τ, so that total waiting costs on $[0, \tau]$ are

$$c(\mathfrak{X}_{0:\tau}) := \sum_{s=0}^{\tau-1} C_{\text{Delay}} \mathbf{1}_{\{I_s^{(2)} > \bar{I}\}} + C_{\text{FA}} \mathbf{1}_{\{I_\tau^{(2)} \leq \bar{I}\}}. \tag{2.3}$$

We will refer to the costs $d(\cdot)$ and $c(\cdot)$ as the immediate cost and the future cost, respectively.

Remark 2 *Note that detection costs are intrinsically defined in terms of the count of infecteds in Pool 2, $I^{(2)}$, which is assumed to be unavailable to the policy-maker. Below we will operationalize* (2.2) *and* (2.3) *by taking conditional expectation with respect to information that is available, see* (4.5)-(4.4).

The aim of outbreak detection is to pinpoint θ, i.e. ideally one takes $\tau = \theta$. However, this is not possible if only partial information is available about $\mathfrak{X}$, specifically about $I_t^{(2)}$. When τ and θ are different, C_{Delay} penalizes the event $\{\tau > \theta\}$, and C_{FA} penalizes $\{\tau < \theta\}$. The cost structure in (2.3) is then a dynamic counterpart of the usual Type-I and Type-II errors in hypothesis testing.

2.1 Detection Problem

Our detection problem is formalized as minimizing the expected future cost over all possible stopping times τ [Poor and Hadjiliadis, 2009], i.e. an optimal stopping problem. Namely, we define the value function V as

$$V(\mathbf{x}_0) := \inf_{\tau \in S} E\left[c(\mathfrak{X}_{0:\tau})|\mathfrak{X}_0 = \mathbf{x}_0\right], \tag{2.4}$$

where $\mathbf{x}_0$ is the initial state. Assuming the infimum in (2.4) is achieved, the dynamic programming principle [Poor and Hadjiliadis, 2009] implies

$$V(\mathbf{x}_0) = \min\left(d(\mathbf{x}_0), E\left[V(\mathfrak{X}_1)|\mathfrak{X}_0 = \mathbf{x}_0\right]\right), \tag{2.5}$$

where the conditional expectation operator is

$$E[V(\mathfrak{X}_1)|\mathfrak{X}_0 = \mathbf{x}_0] = \int V(\mathbf{x}) p_1(\mathbf{x}|\mathbf{x}_0) d\mathbf{x}.$$

The minimum operator in (2.5) corresponds to the idea that it is optimal to declare an outbreak if the immediate cost is smaller than the future cost, i.e. the likelihood of false alarms is dominated by the cost of waiting. The former case is equivalent to the expectation of the value function at time 1 being greater than the immediate cost, and therefore, we may classify the stopping region via

$$\mathfrak{S} := \left\{\mathbf{x} : E\left[V(\mathfrak{X}_1)|\mathfrak{X}_0 = \mathbf{x}\right] - d(\mathbf{x}) > 0\right\}. \tag{2.6}$$

Hence, in terms of the above detection map, our goal is to optimally partition $\mathcal{X} = \mathfrak{S} \cup \mathfrak{C}$ into two regions, such that $\mathfrak{S}$ consists of all initial states $\mathbf{x}_0$ where it is optimal to declare

the epidemic, and $\mathfrak{C}$ is its complement, where it is optimal to wait.

2.2 Reduction to a Model Predictive Control Problem

The characterization in (2.5) is implicit, since it features $V(\cdot)$ on both sides of the expression. Specifically, the value function V corresponds to a fixed point [Bertsekas, 2005] of the functional operator $\mathcal{L}$, defined by $(\mathcal{L}v)(\mathbf{x}) := \min\left(d(\mathbf{x}), E\left[v(\mathfrak{X}_1)|\mathfrak{X}_0 = \mathbf{x}\right]\right)$. To solve for $V(\mathbf{x})$, a basic strategy is then to apply Picard-type fixed-point iterations. In other words, given some initial guess $V^{(0)}(\mathbf{x})$, we build a sequence of approximations via $V^{(k)} := \mathcal{L}V^{(k-1)}$, or explicitly,

$$V^{(k)}(\mathbf{x}_0) = \min\left(d(\mathbf{x}_0), E\left[V^{(k-1)}(\mathfrak{X}_1)|\mathfrak{X}_0 = \mathbf{x}_0\right]\right). \tag{2.7}$$

However, to guarantee the convergence of $V^{(k)}$ does not appear tractable, and the practical performance of (2.7) is very sensitive to the initial guess $V^{(0)}$. To circumvent this challenge, we rely on the concept of model predictive control (also known as receding horizon control). To wit, we introduce an auxiliary parameter t which can be intuitively thought of as *forward time*. The value functions $V(t,\cdot)$ and detection maps $\mathfrak{S}_t$ are now also indexed by t. We start with the trivial initial condition $V(0,\mathbf{x}) := d(\mathbf{x})$, which corresponds to $\mathfrak{S}_0 \equiv \mathcal{X}$. Next, mimicking the classical dynamic programming on finite horizon, we define

$$V(t,\mathbf{x}_0) := \min\left(d(\mathbf{x}_0), E\left[V(t-1,\mathfrak{X}_1)|\mathfrak{X}_0 = \mathbf{x}_0\right]\right), \quad t = 1,2,\ldots. \tag{2.8}$$

Define the Q-value, also known as costs-to-go and current costs by

$$q(t, \mathbf{x}) := E[V(t-1, \mathfrak{X}_1)|\mathfrak{X}_0 = \mathbf{x}] - d(\mathbf{x}). \tag{2.9}$$

Then the stopping set at iteration t is

$$\mathfrak{S}_t := \{\mathbf{x}_0 \in \mathcal{X} : q(t, \mathbf{x}_0) > 0\}, \quad t = 1, 2, \ldots. \tag{2.10}$$

We may "unroll" the expectation encoded in $V(t-1, \mathfrak{X}_1)$ to write

$$E\left[V(t-1, \mathfrak{X}_1)|\mathfrak{X}_0 = \mathbf{x}_0\right] = E\left[c(\mathfrak{X}_{0:\tau^{(t)}})|\mathfrak{X}_0 = \mathbf{x}_0\right], \tag{2.11}$$

where $\tau^{(t)} = \min\{s \geq 1 : \mathfrak{X}_s \in \mathfrak{S}_{t-s}\}$. This justifies the interpretation of $q(t, \cdot)$ as costs-to-go since $c(\mathfrak{X}_{0:\tau^{(t)}})$ are indeed the future costs associated with *not* stopping immediately.

Figure 2.1 illustrates the first step of the recursion (2.8) at $t = 1$. In the plot we compare

$$E\left[V(0, \mathfrak{X}_1)|\mathfrak{X}_0 = \mathbf{x}\right] = E[d(\mathfrak{X}_1)|\mathfrak{X}_0 = \mathbf{x}]$$

against $d(\mathbf{x})$. As discussed, the epidemic is announced when $E\left[V(t-1, \mathfrak{X}_1)|\mathfrak{X}_0 = \mathbf{x}\right] > d(\mathbf{x})$ (the right side of the plot). In the opposite case (the left side of the plot), the optimal decision is to wait. As shown by the Figure 2.1, the structure of the decision map is driven by the regions where these two quantities are equal to each other, which corresponds to the detection *boundary*,

$$\partial\mathfrak{S}_t := \{\mathbf{x} : E\left[V(t-1, \mathfrak{X}_1)|\mathfrak{X}_0 = \mathbf{x}\right] = d(\mathbf{x})\}. \tag{2.12}$$

The stopping region $\mathfrak{S}_t$ is our detection rule at iteration step t, which minimizes our future expected costs. Thus, it can be characterized as the optimal detection rule among

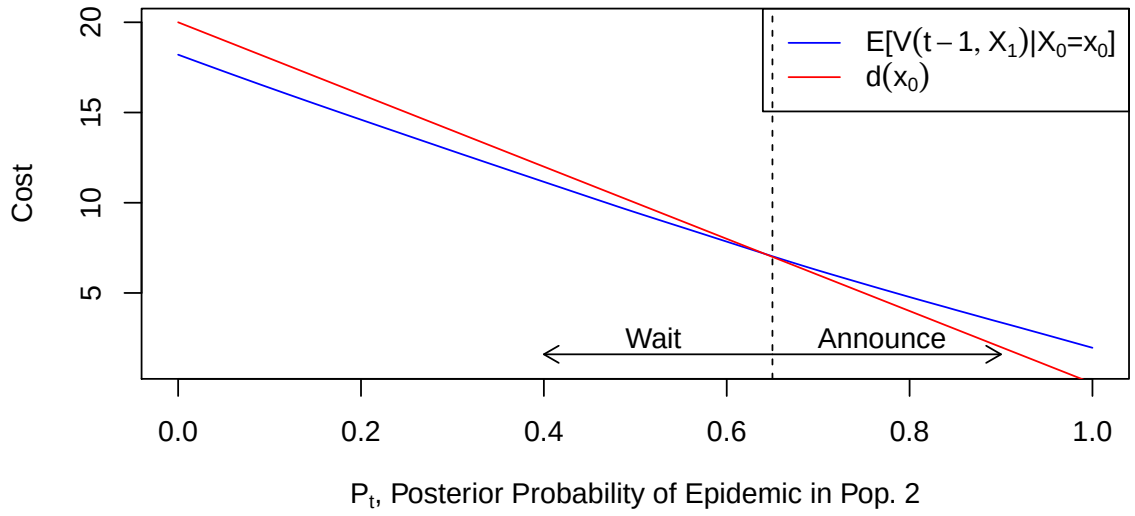


Figure 2.1: Detection strategy at iteration $t = 1$. The example is based on the model of Section 6.1, with parameters in Table 6.1. In the plot, the state of Pool 1 is held fixed at $S_0^{(1)} = 1990,\ I_0^{(1)} = 10$.

all strategies in $\mathcal{S}^{(t)} = \{\tau \in \mathcal{F} : \tau \leq t\}$ that are upper-bounded by t (by construction, $\tau^{(t)} \leq t$). As $t \to \infty$, we have that the set of admissible rules expands $\mathcal{S}^{(t)} \nearrow \mathcal{S}$, and hence we expect that $\mathfrak{S}_t \to \mathfrak{S}$ and $V(t, \mathbf{x}) \to V(\mathbf{x})$. Intuitively, for large t, the recursively defined (2.8) converges to a stationary case that ought to be the fixed point defining $V(x)$ in (2.4). Such convergence is illustrated in Figure 2.2, where we trace the boundaries $\partial\hat{\mathfrak{S}}_t$ for $t = 1, \ldots, 20$. Convergence takes hold after about 15 iterations and suggests that $\hat{\mathfrak{S}}_{20} \simeq \mathfrak{S}$; this is what we used for Figures 6.1-6.5 where the boundary of $\hat{\mathfrak{S}}_{20}$ was taken as the final output of the algorithm.

The above convergence can be improved via model predictive control [Nevistic and Primbs, 1996] which applies the fixed detection map $\hat{\mathfrak{S}}^{(k)}$, rather than the time-dependent $\hat{\mathfrak{S}}_t$ at each step. Model predictive control simplifies this feature with a time-invariant

rule that simply utilizes $\hat{\mathfrak{S}}_{t-1}$ (that we relabel as $\hat{\mathfrak{S}}^{(t-1)}$ for typographical distinction). Indeed, as Figure 2.2 shows, the early maps $\hat{\mathfrak{S}}_1, \hat{\mathfrak{S}}_2, \ldots$, are not as accurate as $\hat{\mathfrak{S}}_{t-1}$ for t large, so it makes sense to completely "forget" them and rely just on the last iteration step. Accordingly, we implement a blend of (2.7) and (2.8) by first using (5.1) over $t = 1, 2, \ldots, t^*$ and then switching to a receding-horizon rule which is defined as

$$\tau_t^{MPC} := \min\{s \geq 1 : \mathfrak{X}_s \in \hat{\mathfrak{S}}_{t-1}\}, \qquad t = t^*, t^* + 1, \ldots. \tag{2.13}$$

The above MPC iterations are terminated once $\hat{q}(t, \mathbf{x})$ and $\hat{q}(t+1, \mathbf{x})$ do not change much, namely $\|\hat{q}(t, \cdot) - \hat{q}(t+1, \cdot)\|_{L^\infty} < Tol$ for a specified tolerance level Tol.

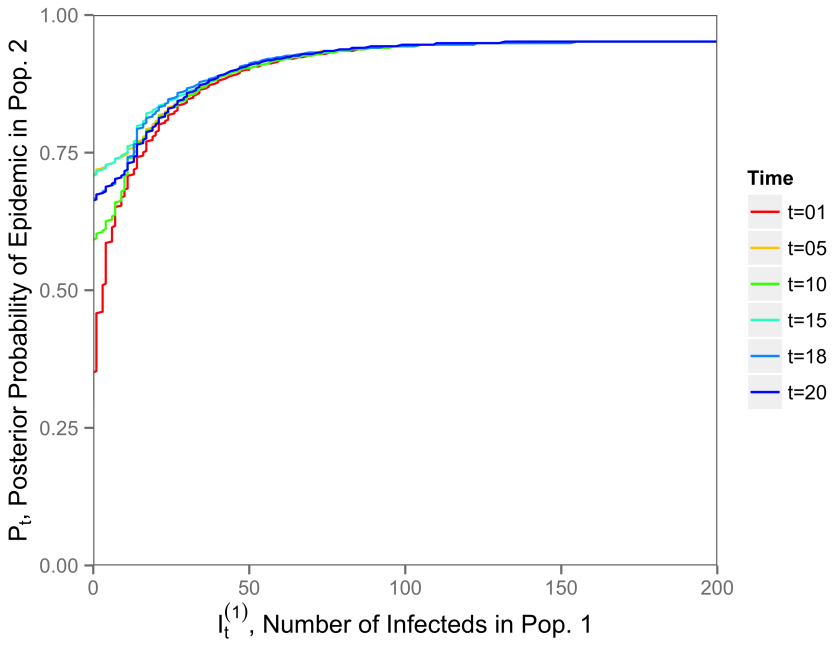


Figure 2.2: Convergence of the detection boundaries $\partial\hat{\mathfrak{S}}_t$ over $t = 1$ to $t = 20$ for the 2-D LP detection rule from Section 6.1.

Thus, we derived a sequence of detection maps given in $\hat{\mathfrak{S}}_t$ for $t = 1, 2, \ldots$, defined recursively in (2.10), which converge to the true $\mathfrak{S}$. In Chapter 5 we empirically estimate each $\hat{\mathfrak{S}}_t$ for $t = 1, 2, \ldots$. In the next chapter we discuss Stochastic Epidemic models,

which leads to the definition of our state space model $\mathfrak{X}_t$ in Chapter 4.

Chapter 3

Stochastic Epidemic Models

Stochastic Epidemic models are used to describe the spread of viral and bacterial infectious diseases with subject-to-subject transmissions. The examples of such diseases are common cold, influenza, measles, chickenpox, rubella, etc. The spread of the diseases depends strongly on how many susceptible and infectious individuals are located in the area. Therefore we would need to talk about compartments – groups of individuals with the same status in regards to the disease. For example, a compartment of Susceptible individuals is a group of people who are at risk of getting a disease. Another important compartment is Infectious – a group of individuals who can transmit the disease. Thus a whole population of a city/state/country can be divided into compartments so that any individual is a member of just one compartment.

Different diseases have their own compartmental models. Some diseases (infuenza, chickenpox, measles, rubella) have full-immunity after a person has been infected and recovered from them [Brauer, 2008]. Therefore, only three compartments will be used for this type of the disease: Susceptible - Infected - Recovered (SIR). There are other modifications: if an infected person does not get an immunity from the disease (such as tuberculosis, meningitis, gonorrhea, sexually transmitted diseases) [Brauer, 2008, Allen

et al., 2008], then after recovering from the disease, this person becomes Susceptible again, and the epidemic spread follows a Susceptible-Infected-Susceptible (SIS) model. Or if a person may get a temporary immunity to the disease (for example, syphilis [Grassly et al., 2005]), then it is an Susceptible - Infected - Recovered - Susceptible (SIRS) model. Some childhood diseases (measles, rubella, or chickenpox [Keeling and Rohani, 2011]) can be modeled with a latent state: a person is infected with the virus, but can not infect others for a certain period of time. Therefore, an Exposed compartment is used for these individuals, and the compartmental model becomes Susceptible - Exposed - Infected - Recovered (SEIR).

The basic reproduction number, which is denoted by R_0, is defined as an expected number of secondary infections caused by one infectious individual in a susceptible population. R_0 determines if an epidemic occurs. If $R_0 < 1$, the infection likely dies out, while if $R_0 > 1$, we expect an epidemic. Note that we defined an epidemic as if the number of infected individuals crosses a specified threshold of $\bar{I}$.

In this thesis we focus on two-population models assuming that the epidemic begins in Pool 1 and may subsequently spread to Pool 2, where it is to be detected. For example, we can consider all people in state of California being Pool 1 and all people in the state of Arizona being Pool 2, or the Los Angeles population being Pool 1 and the New York population being Pool 2. Accordingly, we will be fusing information from Pool 1 and Pool 2 to identify the onset of an epidemic in Pool 2. The parameters are assumed to be known and are a function of the modeled disease family (e.g. influenza or dengue fever), the demographics and public health characteristics of the populations, and the travel patterns across pools.

Note that while the discussion up to this point works equally well for stochastic and Deterministic models, in this thesis we focus our attention on stochastic models. Stochastic models provide variability of the model outcomes, which mimic the possible

outcomes that could occur in real life. Deterministic models with the same starting conditions will always give the same results which is not possible in reality.

In Section 3.1.1 we will define simple models for one population such as SIR, SIR with vital dynamics, SEIR, and SEIR with vital dynamics, respectively. Then in Section 3.2 we will define two-population models and show how two populations can interact between each other. In Section 3.3 we will demonstrate UK Measles dataset and discuss the features of SEIR Model in regards to this particular dataset.

3.1 One Population Models

3.1.1 SIR/SEIR Model

A basic stochastic compartmental model is the Susceptible-Infected-Recovered (SIR) model which consists of three eponymous compartments: Susceptible, Infectious, and Recovered. Susceptible individuals are the ones who have not experienced the disease yet. Interaction between an infected and a susceptible individual can lead to an infection. Thus, this kind of interaction stochastically generates new infecteds who, in turn, can further infect other susceptible individuals. After some time an infected individual recovers and becomes immune, i.e. becomes a Recovered: he/she can neither infect others nor get infected.

A Susceptible-Exposed-Infected-Recovered (SEIR) model defined by He et al. [2010], Brauer [2008], Hethcote [2000], Dukic et al. [2012], Smith et al. [2001] assumes the division of population into 4 compartments: S, susceptible to infectious individuals; E, exposed (infected, but not yet infectious); I, infectious; R, recovered. So the main difference from the SIR model defined above is the latent compartment E in which individuals are already infected (they might be experiencing symptoms) but cannot infect others yet. If

the exposed period is short, it is often neglected in modeling. A longer exposed period might lead to different model predictions, and we can think of the exposed compartment as an estimate of infectious individuals tomorrow.

For describing outbreak dynamics it is more convenient to work with continuous-time dynamical systems, but later, we will define its discretized version. The overall epidemic state at epoch $t \in \mathbb{R}_+$ of pool k is denoted by $\mathbf{X}_t^{(k)} = \{S_t^{(k)}, E_t^{(k)}, I_t^{(k)}, R_t^{(k)}\}$, where $S_t^{(k)}$, $E_t^{(k)}$, $I_t^{(k)}$ and $R_t^{(k)}$ are the counts of susceptible, exposed, infectious and recovered individuals in the population k, respectively. If we assume that the pool size of our population is fixed at $M^{(k)} = S_t^{(k)} + E_t^{(k)} + I_t^{(k)} + R_t^{(k)}$, we can omit further mention of $R_t^{(k)}$ since $R_t^{(k)} = M^{(k)} - S_t^{(k)} - E_t^{(k)} - I_t^{(k)}$. Note that if we have an SIR model, we can omit the exposed compartment $E_t^{(k)}$. The unit for time t is usually taken as a week.

The continuous evolution of the state process $\{S_t^{(k)}, E_t^{(k)}, I_t^{(k)}\} \in \{(s, e, i) : s+e+i \leq M^{(k)}\}$ is described using a Markov chain or stochastic kinetic system language. Namely, the epidemic state is piecewise constant in time. Next, there are three possible transitions described by the following reaction channels [He et al., 2010, Wilkinson, 2006, Allen et al., 2008, Andersson and Britton, 2000]:

$$\left\{ \begin{array}{rrlll} \text{Infection} & S^{(k)} + I^{(k)} & \to E^{(k)} + I^{(k)} & \text{w/rate} & \mu_{SI}^{(k)}(t)(I^{(k)} + \mathbf{i}^{(k)})^{\alpha^{(k)}} \frac{S^{(k)}}{M^{(k)}} \\ \text{Contagiousness} & E^{(k)} & \to I^{(k)} & \text{w/rate} & \mu_{EI}^{(k)} E^{(k)} \\ \text{Recovery} & I^{(k)} & \to \emptyset & \text{w/rate} & \mu_{IR}^{(k)} I^{(k)} \end{array} \right\} \quad (3.1)$$

The graphical interpretation of these reaction channels is given on Figure 3.1.

The first transition represents an infection of a susceptible individual by an infectious individual. This transition happens at rate $\mu_{SI}^{(k)}(t)(I_t^{(k)} + \mathbf{i}^{(k)})^{\alpha^{(k)}} \frac{S_t^{(k)}}{M^{(k)}}$, where $\mu_{SI}^{(k)}(t)$ is the contact rate of infected and susceptible individuals within meta-population k at time t, $\mathbf{i}$ is the mean number of all infectives visiting the population at any given time, and α is the mixing parameter, with $\alpha = 1$ corresponding to homogeneous mixing [Bjørnstad

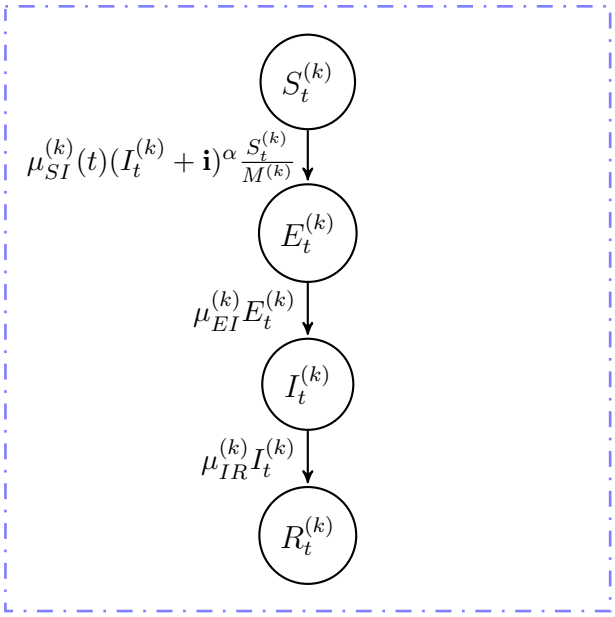


Figure 3.1: Stochastic SEIR epidemic model

et al., 2002, Liu et al., 1986]. The derivation of this rate is provided in Brauer [2008].

The contact rate $\mu_{SI}^{(k)}(t)$ is time-dependent because the contact rates are related to the weather/temperature parameters (in case of common cold/influenza), school terms (in case of childhood diseases), etc. For example, papers Liu et al. [1986], Schenzle [1984], Keeling and Grenfell [2002], Conlan and Grenfell [2007], He et al. [2010] study the spread of measles in UK, and they found that the contact rate $\mu_{SI}^{(k)}(t)$ reflects the pattern of British school terms and holidays. He et al. [2010] takes the contact rate to be:

$$\mu_{SI}^{(k)}(t) = \begin{cases} \left(1+2\left(1-p\right)a^{(k)}\right)\bar{\mu}_{SI}^{(k)}, & \text{during school term,} \\ \left(1-2pa^{(k)}\right)\bar{\mu}_{SI}^{(k)}, & \text{during school holiday,} \end{cases} \tag{3.2}$$

where $p = 0.759$ is the proportion of the year taken up by the school term, $\bar{\mu}_{SI}^{(k)}$ the mean contact rate, and $a^{(k)}$ the relative effect of school holidays on transmission in population k.

The second transition in Equation (3.1) corresponds to the gained ability to infect others, i.e. become infectious/contagious. The rate of this transition is $\mu_{EI}^{(k)} E_t^{(k)}$, where $1/\mu_{EI}^{(k)}$ is the mean exposed period. So in case of SIR model, we can take $\mu_{EI}^{(k)} = \infty$ and therefore there an instant transition to an Infectious compartment.

The third transition in (3.1) represents recovery and subsequent immunity of an infected individual. The rate of transition is $\mu_{IR}^{(k)} I_t^{(k)}$, where $\mu_{IR}^{(k)}$ is a recovery rate. This can be interpreted as individuals staying infected for an Exponentially distributed time with mean $1/\mu_{IR}^{(k)}$. We assume that the recovery rate is constant over time.

The time until each transition is Exponential with the corresponding rate, and the time until next transition is Exponential with rate – sum of the rates in (3.1).

For an SIR model with homogeneous mixing $\alpha = 1$ the basic reproduction number [Andersson and Britton, 2000] can be found as

$$R_0(t) = \frac{\mu_{SI}^{(k)}(t)}{\mu_{IR}^{(k)}}. \tag{3.3}$$

Therefore for an epidemic to start, we need the infection rate $\mu_{SI}^{(k)}(t)$ to be greater than the recovery rate $\mu_{IR}^{(k)}$.

The discrete time SEIR model $\{S_n^{(k)}, E_n^{(k)}, I_n^{(k)}\},\ n = 1, 2, \ldots$ is a Markov chain with finite state space. Suppose we have a small time interval Δt such that only one transition (either infection or recovery) happens in this time step. If we denote the new exposed, infectious , and recovered individuals at time step n in population k by $\Delta E_n^{(k)}$, $\Delta I_n^{(k)}$ and $\Delta R_n^{(k)}$, respectively, where $n = 0, 1, 2, \ldots$, then the discrete time SIR has a joint probability function $P(S_n^{(k)} = s, E_n^{(k)} = e, I_n^{(k)} = i)$, where $s, e, i = 0, 1, 2, \ldots, M^{(k)}$ and

$s + e + i \leq M$, as well as having the transition probabilities

$$P(S_{n+1}^{(k)} = s-1, E_n^{(k)} = e+1, I_{n+1}^{(k)} = i | S_n^{(k)} = s, E_n^{(k)} = e, I_n^{(k)} = i) = \mu_{SI}(n\Delta t) i \frac{s}{M^{(k)}} \Delta t, \tag{3.4}$$

$$P(S_{n+1}^{(k)} = s, E_n^{(k)} = e-1, I_{n+1}^{(k)} = i+1 | S_n^{(k)} = s, E_n^{(k)} = e, I_n^{(k)} = i) = \mu_{EI} e \Delta t, \tag{3.5}$$

$$P(S_{n+1}^{(k)} = s, E_n^{(k)} = e, I_{n+1}^{(k)} = i-1 | S_n^{(k)} = s, E_n^{(k)} = e, I_n^{(k)} = i) = \mu_{IR} i \Delta t, \tag{3.6}$$

$$P(S_{n+1}^{(k)} = s, E_n^{(k)} = e, I_{n+1}^{(k)} = i | S_n^{(k)} = s, E_n^{(k)} = e, I_n^{(k)} = i) = 1 - \left(\mu_{SI}(n\Delta t) \frac{is}{M^{(k)}} + \mu_{EI} e + \mu_{IR} i \right) \Delta t. \tag{3.7}$$

We also have the conditions $\Delta t(\mu_{SI}(n\Delta t)\frac{is}{M^{(k)}} + \mu_{EI} i + \mu_{IR} i) \leq 1$ for all $s, i = 0, 1, 2, \ldots, M^{(k)}$ and $s+e+i \leq M^{(k)}$ [Allen and Burgin, 2000] to guarantee that the transitions probabilities are valid probabilities and the compartment sizes stay greater than zero.

The above transition probabilities approximate the transition probabilities of a continuous-time Markov jump process, where the changes in the compartments follow a Poisson process, and the time between jumps is given by an exponential distribution with mean $1/(\mu_{SI}(n\Delta t) i \frac{s}{M^{(k)}} + \mu_{EI} i + \mu_{IR} i)$ [Allen and Burgin, 2000].

If we denote the new exposed, infectious individuals and recovered individuals at time step n, where $n = 0, 1, 2, \ldots$, in population k by $\Delta E_n^{(k)}$, $\Delta I_n^{(k)}$ and $\Delta R_n^{(k)}$, respectively, then for a time change of 1 week:

$$\Delta E_n^{(k)} \sim Poisson \left(\mu_{SI}^{(k)}(n) (I_n^{(k)} + \mathbf{i}^{(k)})^{\alpha^{(k)}} \frac{S_n^{(k)}}{M_n^{(k)}} \right), \tag{3.8}$$

$$\Delta I_n^{(k)} \sim Poisson \left(\mu_{EI}^{(k)} E_n^{(k)} \right) \tag{3.9}$$

$$\Delta R_n^{(k)} \sim Poisson \left(\mu_{IR}^{(k)} I_n^{(k)} \right) \tag{3.10}$$

where "$\sim Poisson\,(\cdot)$" means 'distributed as a poisson random variable with the rate' inside the parenthesis. All Poisson random variables are assumed to be independent of

the infectious periods as well as independent of each other because the number of new recovered individuals is independent of the number of people who got sick. Then using equations (3.8), (3.9), and (3.10) we can construct the state process $\{S_n^{(k)}, E_n^{(k)}, I_n^{(k)}\}$ to be:

$$S_{n+1}^{(k)} = S_n^{(k)} - \Delta E_n^{(k)}, \tag{3.11}$$

$$E_{n+1}^{(k)} = E_n^{(k)} + \Delta E_n^{(k)} - \Delta I_n^{(k)}, \tag{3.12}$$

$$I_{n+1}^{(k)} = I_n^{(k)} + \Delta I_n^{(k)} - \Delta R_n^{(k)}, \tag{3.13}$$

with $S_0^{(k)} \geq 0$, $E_0^{(k)} \geq 0$, $I_0^{(k)} \geq 0$, and $S_n^{(k)} + E_n^{(k)} + I_n^{(k)} \leq M^{(k)}$ for all $n = 0, 1, 2, \ldots$. It is possible to encounter different problems: compartments becoming too big and/or negative. To resolve this we overwrite the negative values with zeros.

Since our biosurveillance depends on the number of people who actually reported their disease (i.e. came to clinic to be officially diagnosed) we define $C^{obs(k)}$ as the number of observed (i.e. reported) infected cases.If the non-reporting probability of new infecteds $\Delta I_n^{(k)}$ is ρ, then the distribution of $C^{obs(k)}$ given $\Delta I_n^{(k)}$ is a binomial distribution with parameters $\Delta I_n^{(k)}$ and $1 - \rho$ and therefore it can be shown that

$$\Delta C_n^{obs(k)} \sim Poisson((1-\rho)\Delta I_n^{(k)}). \tag{3.14}$$

3.1.2 SEIR Model with Vital Dynamics

The definition of the SEIR model given in Subsection 3.1.1 assumed a constant population size. This model is good if the epidemic happens quickly and therefore there are no significant changes in the population size. However if the epidemic takes years to develop then it is critical to observe the changes in the population size. Thus we define an SEIR

model with vital dynamics, i.e. with births and deaths [Brauer, 2008, Bjørnstad et al., 2002, Allen et al., 2008, Andersson and Britton, 2000]. Suppose that the individuals are born at rate $\mu_B^{(k)} \cdot M_t^{(k)}$, and each of them has an exponentially distributed lifetime with intensity $\mu_D^{(k)}$ [Andersson and Britton, 2000]. If $\mu_B^{(k)} = \mu_D^{(k)}$, the population size will oscillate around the quantity $M_0^{(k)}$ [Andersson and Britton, 2000].

Suppose an individual can be born immune or become immune shortly after birth to the disease with probability p_{immune}. For example, a newborn can be vaccinated against some diseases, while for some diseases like Whooping Cough, a newborn is protected at birth due to mandatory vaccination of pregnant women. Then if a born individual is immune, then she/he would be placed in a "Recovered" compartment, else she/he would be placed in a "Susceptible" compartment. An individual may die as a Susceptible, Exposed, Infected, or Recovered.

The continuous evolution of the state process $\{S_t^{(k)}, E_t^{(k)}, I_t^{(k)}, R_t^{(k)}\} \in \{(s,i,r) : s + i + r = M_t^{(k)}\}$ is described with a Markov chain with nine possible transitions described by the following reaction channels:

$$\left\{\begin{array}{rrll}
\text{Birth of Susceptible} & \emptyset & \to S^{(k)} & \text{w/rate } (1-p_{immune})\mu_B^{(k)} M^{(k)} \\
\text{Birth of Recovered} & \emptyset & \to R^{(k)} & \text{w/rate } p_{immune}\ \mu_B^{(k)} M^{(k)} \\
\text{Infection} & S^{(k)} + I^{(k)} & \to E^{(k)} + I^{(k)} & \text{w/rate } \mu_{SI}^{(k)}(t)(I^{(k)} + \mathbf{i})^{\alpha} \frac{S^{(k)}}{M^{(k)}} \\
\text{Contagiousness} & E^{(k)} & \to I^{(k)} & \text{w/rate } \mu_{EI}^{(k)} E^{(k)} \\
\text{Recovery} & I^{(k)} & \to R^{(k)} & \text{w/rate } \mu_{IR}^{(k)} I^{(k)} \\
\text{Death of Susceptible} & S^{(k)} & \to \emptyset & \text{w/rate } \mu_D^{(k)} S^{(k)} \\
\text{Death of Exposed} & E^{(k)} & \to \emptyset & \text{w/rate } \mu_D^{(k)} E^{(k)} \\
\text{Death of Infected} & I^{(k)} & \to \emptyset & \text{w/rate } \mu_D^{(k)} I^{(k)} \\
\text{Death of Recovered} & R^{(k)} & \to \emptyset & \text{w/rate } \mu_D^{(k)} R^{(k)}
\end{array}\right\} \tag{3.15}$$

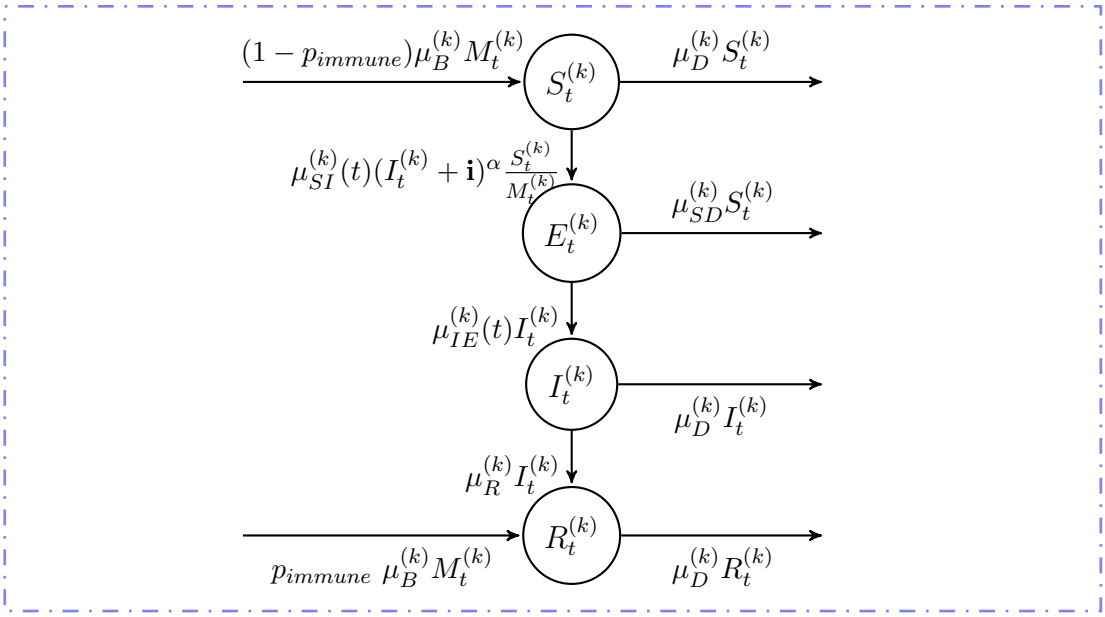


Figure 3.2: Stochastic SEIR epidemic model with vital dynamics in population k.

The graphical interpretation of these reaction channels is given on Figure 3.2.

For the discretized version we denote the newborn individuals with and without immunity as $\Delta(BS)_n^{(k)}$ and $\Delta(BR)_n^{(k)}$, respectively, and new deceased individuals from compartments "Susceptible", "Exposed", "Infectious", "Recovered" at time step n, where $n = 0, 1, 2, \ldots$, in population k by $\Delta(DS)_n^{(k)}$, $\Delta(DI)_n^{(k)}$ and $\Delta(DR)_n^{(k)}$,respectively. We then have

$$\Delta(BS)_n^{(k)} \sim Poisson\left((1-p_{immune})\mu_B^{(k)}M_n^{(k)}\right), \tag{3.16}$$

$$\Delta(BR)_n^{(k)} \sim Poisson\left(p_{immune}\ \mu_B^{(k)}M_n^{(k)}\right), \tag{3.17}$$

$$\Delta(DS)_n^{(k)} \sim Poisson\left(\mu_D^{(k)}S_n^{(k)}\right), \tag{3.18}$$

$$\Delta(DE)_n^{(k)} \sim Poisson\left(\mu_D^{(k)}E_n^{(k)}\right), \tag{3.19}$$

$$\Delta(DI)_n^{(k)} \sim Poisson\left(\mu_D^{(k)}I_n^{(k)}\right), \tag{3.20}$$

$$\Delta(DR)_n^{(k)} \sim Poisson\left(\mu_D^{(k)}R_n^{(k)}\right). \tag{3.21}$$

Using equations (3.16), (3.17), (3.18), (3.9), (3.20), and (3.21) as well as previously defined equations (3.8), (3.9), and (3.10), by construction, the state process $\{S_n^{(k)}, {}_n^{(k)}, I_n^{(k)}, R_n^{(k)}\}$ evolves the following way:

$$S_{n+1}^{(k)} = S_n^{(k)} - \Delta E_n^{(k)} + \Delta(BS)_n^{(k)} - \Delta(DS)_n^{(k)} \tag{3.22}$$

$$E_{n+1}^{(k)} = E_n^{(k)} + \Delta E_n^{(k)} - \Delta I_n^{(k)} - \Delta(DE)_n^{(k)}, \tag{3.23}$$

$$I_{n+1}^{(k)} = I_n^{(k)} + \Delta I_n^{(k)} - \Delta R_n^{(k)} - \Delta(DI)_n^{(k)}, \tag{3.24}$$

$$R_{n+1}^{(k)} = R_n^{(k)} + \Delta R_n^{(k)} + \Delta(BR)_n^{(k)} - \Delta(DR)_n^{(k)} \tag{3.25}$$

with $S_0^{(k)} \geq 0,\ I_0^{(k)} \geq 0$, and $R_0^{(k)} \geq 0$ for all $n = 0, 1, 2, \ldots$.

For an SEIR model with vital dynamics the basic reproduction number [Smith et al., 2001, Hethcote, 2000] can be found as

$$R_0(t) = \frac{\mu_{SI}^{(k)}(t)\mu_{EI}^{(k)}}{(\mu_{EI}^{(k)} + \mu_D^{(k)})(\mu_D^{(k)} + \mu_{IR}^{(k)})}, \tag{3.26}$$

assuming homogeneous mixing.

3.2 Two Population Models

As in Chapter 6 of Andersson and Britton [2000], we first recall the multi-type stochastic SIR model in continuous time. The overall epidemic state at epoch t is denoted by $\mathbf{X}_t = \{\mathbf{X}_t^{(1)}, \mathbf{X}_t^{(2)}\}$, where each $\mathbf{X}_t^{(k)}$ denotes the epidemic state governed by a specific compartmental model in population $k = 1, 2$. The choice of compartmental model for each population is determined by the specifics of the disease as well as the specifics of the meta-population itself.

A meta-population in this case is two (or more) spatially separated populations which

interact between each other (as we discussed in the beginning of this Chapter 3, it could be two different states in USA or two different cities). The amount of interaction depends on the population sizes as well as the amount of trade and other commercial activities, immigration, number of tourists, etc. Each of these interactions can spread the infection from one population to another.

The 'transmission' transition is added to the dynamics of the model: the infection of a susceptible individual from pool k by an infected individual from a different pool k'. The frequency of such infections is $\mu_{SI}^{(k,k')}(t) I_t^{(k')} \frac{S_t^{(k)}}{M^{(k)}}$, where $1 \leq k, k' \leq K$ and $\mu_{SI}^{(k',k)}(t)$ is the contact rate of infected individuals from population k' with a susceptible individual from population k. Since contacts between individuals from different populations are less frequent, $\mu_{SI}^{(k,k')}(t) \ll \mu_{SI}^{(k')}(t)$ for any t. To reduce the number of parameters, we thus assume that cross-population interactions occur at rate $\mu_{SI}^{(k,k')}(t) \equiv \gamma^{(k,k')} \mu_{SI}^{(k')}(t)$, where $\gamma^{(k,k')}$ is the proportion of "travelers" in each pool adjusted to the population size:

$$\gamma^{(k,k')} = \begin{cases} \gamma, & \text{if } M_t^{(k)} \geq M_t^{(k')}, \\ \gamma \cdot \frac{M_t^{(k)}}{M_t^{(k')}}, & \text{if } M_t^{(k)} < M_t^{(k')}. \end{cases} \tag{3.27}$$

Thus, cross-contacts happen at the fraction γ of the contact rate within one population; a typical range for γ is $[0.01, 0.2]$.

The intuition behind adjusting for population sizes is that a higher proportion of "travelers" from a small city will visit a megapolis rather than the other way around. The travelers from a megapolis will have a bigger variety of cities to visit: they can travel to many cities visiting a particular small city, while most travelers from a small city will have to travel to the megapolis first and then travel to their destination or just stay there. Suppose we have two cities: Metropolis and Smallville, and their populations are 1 million and 10,000 individuals, respectively. Furthermore, suppose travelers comprise 10% of the

population in each city ($\gamma = 0.1$): 100,000 in Metropolis and 1000 in Smallville. The travelers from Smallville will more likely to go to Metropolis, while the travelers from Metropolis will travel to different cities, and only a portion of them will go to Smallville. If we assume that only 10% of Metropolis' travelers land in Smallville, then these 10% or 0.1 is just the ratio of the population of the Smallville to the population of Metropolis.

Figure 3.3 represents the interaction between two populations, where the epidemic spread in the first and second populations is governed by an SEIR model with vital dynamics and an SIR model with vital dynamics, respectively, i.e.
$\mathbf{X}_t^{(1)} = \{S_t^{(1)}, E_t^{(1)}, I_t^{(1)}, R_t^{(1)}\}$ and $\mathbf{X}_t^{(2)} = \{S_t^{(2)}, I_t^{(2)}, R_t^{(2)}\}$. The reaction channels for the first and second populations are provided in (3.15) with an additional "Contagiousness" transition for the second population with its form given in Equation (3.1). The cross-population transmission of the disease occurs using these reaction channels:

$$\left\{ \begin{array}{llll} \text{From Pool 1 to Pool 2} & I^{(1)} + S^{(2)} \to I^{(1)} + E^{(2)} & \text{w/rate} & \mu_{SI}^{(1,2)}(t) I^{(1)} \frac{S^{(2)}}{M^{(2)}} \\ \text{From Pool 2 to Pool 1} & S^{(1)} + I^{(2)} \to E^{(1)} + I^{(2)} & \text{w/rate} & \mu_{SI}^{(2,1)}(t) I^{(2)} \frac{S^{(1)}}{M^{(1)}} \end{array} \right\} \tag{3.28}$$

We also define a Two Population Asymmetrical epidemic model with $\mu_{SI}^{(1,2)} = 0$. This model allows only one-way transmissions of epidemic from Pool 1 to Pool 2, so Pool 2 can not infect Pool 1. So there is only one cross-population transition with the rate $\mu_{SI}^{(2,1)}(t) I_t^{(1)} \frac{S_t^{(2)}}{M_t^{(2)}}$.

Our main goal is to make a decision about the announcement of epidemic in Pool 2 using the information from Pool 1. In this situation, if the population size of Pool 1 is much larger than Pool 2 (for example, $M_t^{(1)} \approx 100 M_t^{(2)}$), and hence the number of infected individuals in Pool 2, $I_t^{(2)}$, is small compared to the the number of infected individuals in Pool 1, $I_t^{(1)}$, then the impact of epidemic transmission from Pool 2 to Pool

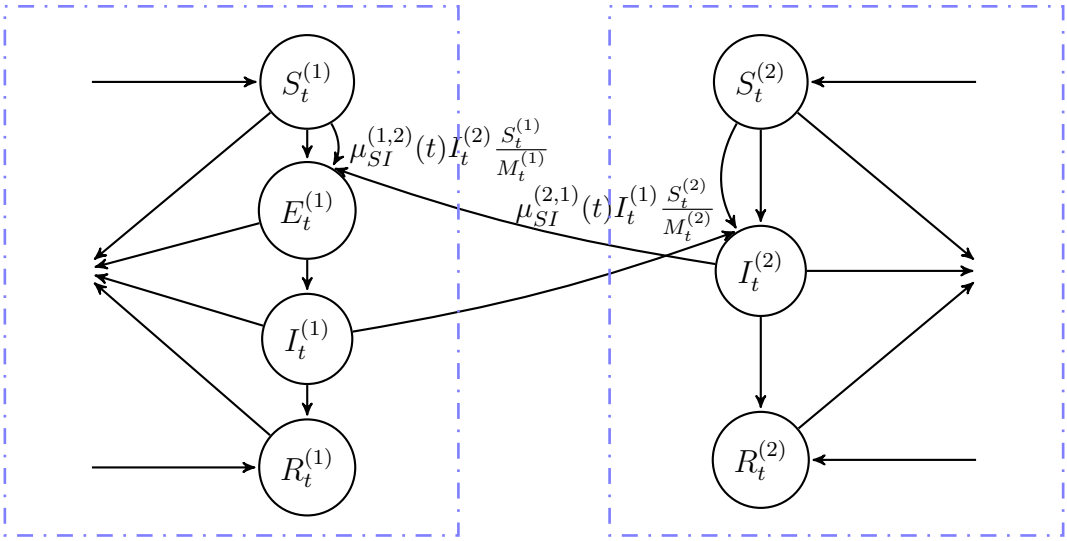


Figure 3.3: Stochastic SEIR-SIR epidemic model with vital dynamics in two populations. The rates of transitions inside each population are given in Equation (3.15) with an additional "Contagiousness" transition for the second population with its form given in Equation (3.1).

1 is negligible and would not have a big effect on $I_t^{(1)}$.

The two population model will work with both continuous-time and discrete-time models inserted in each population's model.

3.3 UK Study Description and SEIR Model

We will illustrate our detection method using a case study of measles transmission dynamics in England and Wales. This study contains weekly case reports from 6 major cities (London, Birmingham, Manchester, Liverpool, Sheffield, Bristol) over the period 1948–1987. See Figure 3.4 for associated time-series plots. The national immunization program began in 1968, therefore the weekly number of cases significantly decreased after 1968.

The coloring of the lines on Figure 3.4 correspond to the school status: whether it is a holiday or regular term. It can be easily seen that the increase in the number of

reported cases happens during the school term and therefore the infection rate at those times is much higher as seen in the definition of the contact rate in Equation (3.2).

The biennial trend of epidemic, which can be seen on Figure 3.4, depends on the number of Susceptibles individuals. The bigger the proportion of Susceptibles in the population, the sooner the epidemic happens. Therefore some cities have an epidemic every year (Liverpool, for example), while others (for example, London) have an epidemic every other year.

We observe moderate-to-strong strength of correlation between the London data and the data from other cities (see Figure 3.5), meaning that the measles outbreak is not local and spread around the country.

He et al. [2010] estimated the parameters of this dataset, and the results are provided in Table 3.1. Note that the contact rate $\bar{\mu}_{SI}$ can be found using Equations (3.2) and (3.26). The rates provided in Table 3.1 are weekly rates, therefore the simulations are weekly. The paper by He et al. [2010] was using the birth rate data, which we were unable to obtain, so we assume that the birth rate is equal to mortality rate $\mu_D(t) = 3.836 \cdot 10^{-4}$ per week and the instantaneous birth rate is equal to

$$\mu_B(t) = \bar{\mu}_B(t) \left(1 - c + 52c\delta(t - t_0 \bmod 52)\right), \tag{3.29}$$

where $\bar{\mu}_B(t) = \mu_D(t) = 3.836{*}{\cdot}10^{-4}$ is the birth rate, the term $\delta(t - t_0 \bmod 52)$ contributes a Dirac delta impulse to the birth rate when t falls on the same calendar week as t_0, and we take $t_0 = 37$ (which corresponds to the first week of September) as a school admission week in England and Wales. Therefore we get an influx of Susceptible individuals every year during the first week of September.

To initialize our simulations we need to provide starting conditions. However since these values are unknown, we provide a guess and then remove a burnout period. So

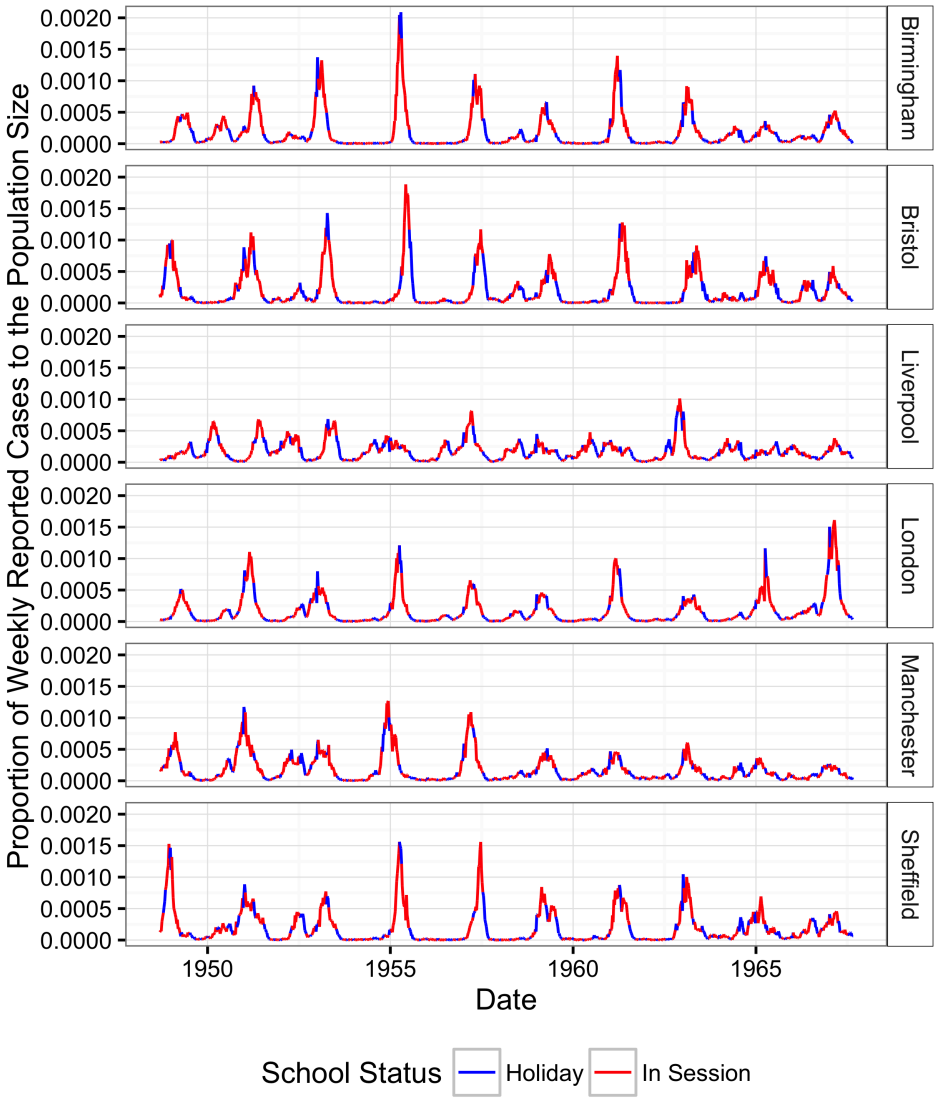


Figure 3.4: Time Series of the Measles Epidemic in UK cities in 1948–1968.

	Birmingham	Sheffield	Manchester	Bristol	Liverpool
Correlation	0.71	0.54	0.52	0.47	0.3
Distance	164	228	262	171	287

Figure 3.5: Correlation of London cases with other cities's cases and distance between London and other cities (in km)

	M	α	R_0	μ_{EI}	μ_{IR}	a	$\mathbf{i}$	c	$1-\rho$
London	3 390	0.98	57	0.554	0.583	0.55	2.90	0.56	0.49
Birmingham	1 118	1.02	35	0.653	0.947	0.31	1.09	0.61	0.56
Liverpool	802	0.98	48	0.947	0.753	0.30	0.26	0.19	0.49
Manchester	704	0.97	33	0.660	1.089	0.29	0.59	0.36	0.55
Sheffield	515	1.02	33	1.042	1.193	0.31	0.85	0.23	0.65
Bristol	443	1.01	27	1.233	1.583	0.20	0.44	0.34	0.63

Table 3.1: Parameter Estimates for UK cities of population size in thousands M, mixing exponent α, reproduction ratio R_0, weekly rates of transition from exposed compartment E to infectious compartment I μ_{EI}, weekly recovery rate μ_{IR}, amplitude of seasonality a, mean number of visiting infectives $\mathbf{i}$, fraction of 'susceptible' individuals enter on the school admission day c, reporting probability $1-\rho$.

as a starting point for our epidemic, we assume that 0.5% of the whole population is in the Susceptible compartment and 0.002% of whole population is in the Exposed and Infectious compartments. We simulated 1000 epidemics for five years in London, removed the first four years from our analysis, and then compared our simulated data to the London data for the 1956-1957 school year (see Figure 3.6). We observe that original London data falls inside our 95% quantile interval of simulated data with exception of short period around July. Therefore, our model is adequate for this data.

We also generate 1000 joint epidemics in London and Bristol for 15 years using Two Population Asymmetric SEIR model with $\gamma = \frac{\mathbf{i}^{(k')}}{\mathbf{i}^{(k)}}\bar{\mu}_{SI}^{(k)} \approx 0.02\bar{\mu}_{SI}^{(k)}$, then remove the first 5 years from our analysis, and compare them to the London-Bristol data for the school years 1954-1958. Figure 3.6 illustrates this comparison and we can observe that 53.8% of the time the London counts are within the 95% quantile regions and 91.8% of the time the Bristol counts are within the 95% quantile regions.

If we look closely at the shapes of the epidemics in Figure 3.6 we see that the epidemics during school years 1954-1955 and 1956-1957 are significantly larger than epidemics during school years 1955-1956 and 1957-1958. The significance of epidemic depends on the number of susceptibles. Figure 3.7 shows that there were significant epidemics during

the 1st, 2nd, and 4th years with no significant epidemic during the 3rd year due to an insufficient number of Susceptible individuals in the 3rd year.

Both plots on the Figure 3.6 show a possible peak of epidemic in late Winter/Early Spring, however only a bivariate model (bottom plot on the Figure 3.6) depicts a possible peak of epidemic also during late Spring. Therefore nearby cities do affect the epidemics, and therefore they should be included in the model.

Thus, we defined Two-Population models in Section and showed that the simulations from our models can mimic the real behavior of epidemics. However this models only work when we observe all information about both populations. In the following chapter we discuss models, where we only observe partial information or no information about the second population.

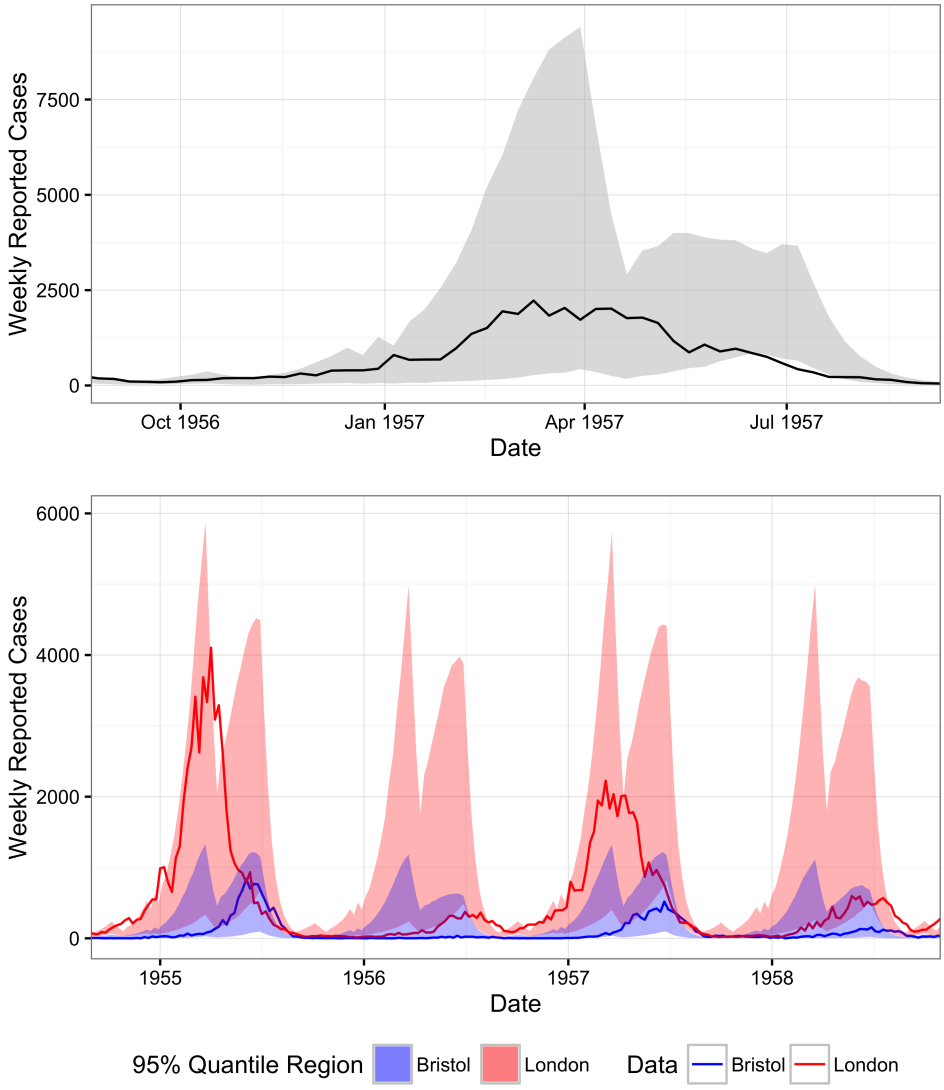


Figure 3.6: Top: A 95% quantile interval of simulated Weekly Reported cases (shaded region) vs. original Weekly Reported Cases (black line) in London in the school year 1956-1957. Bottom: A 95% quantile interval of simulated Weekly Reported cases (shaded regions) vs. original Weekly Reported Cases (lines) in London (red) and Bristol (blue) in the school years 1954 – 1958 .

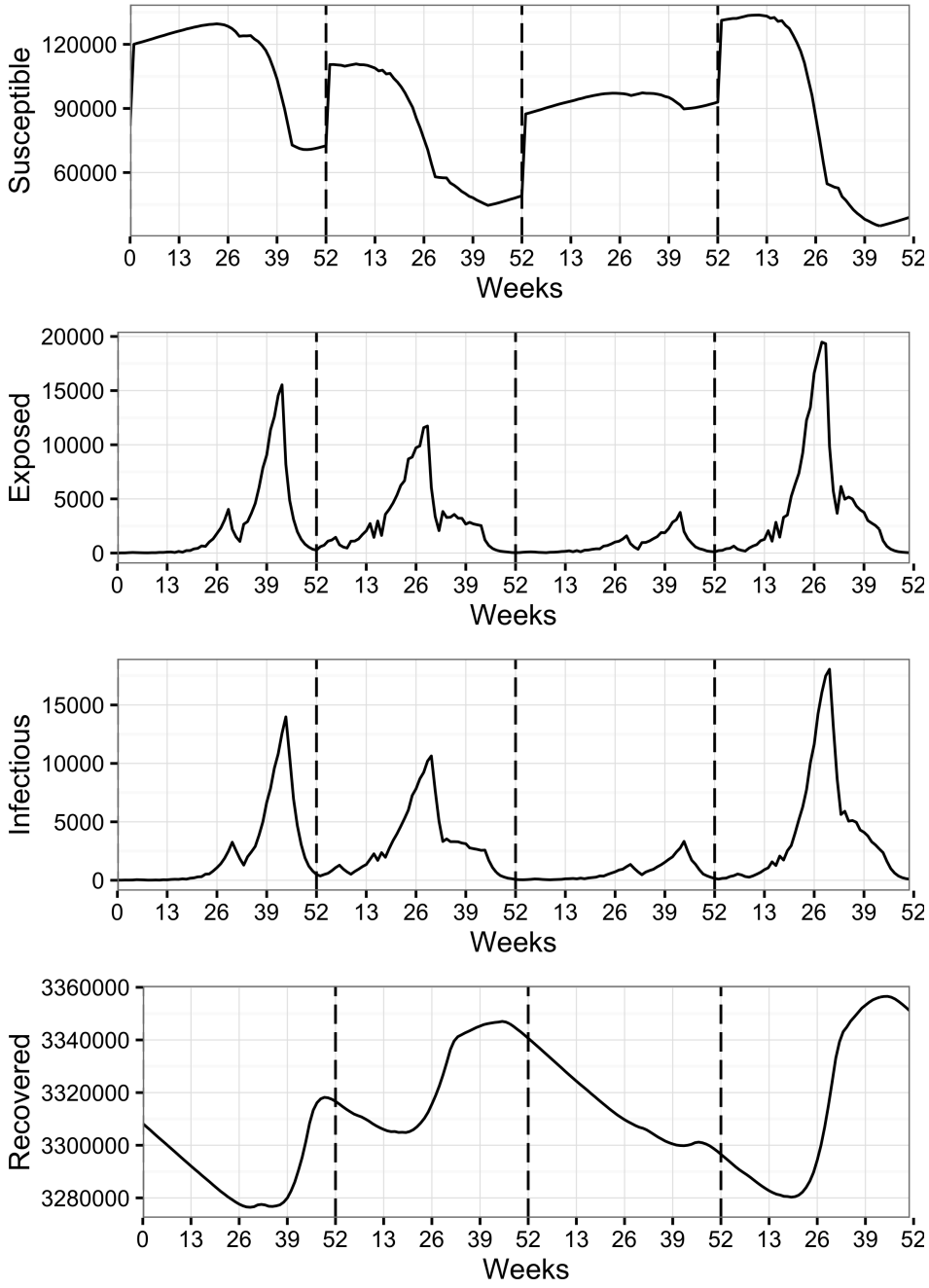


Figure 3.7: Simulated counts of Susceptible, Exposed, Infectious and Recovered individuals over 4 years. Week 0 is the start of the School year. The dashed lines represent the beginning of 2nd, 3rd, and 4th years.

Chapter 4

Reduced Models

Under full observations the detection problem (2.4) would be trivial, since one can directly track $I_t^{(2)}$ and declare an outbreak as soon as there any infecteds in the second pool. Realistically, however, $I^{(2)}$ is not observed. Some of the reasons include mis-diagnoses among infecteds, patients not seeking care, false positives, mis-reporting or lack of reporting of epidemiological data, etc. Consequently, we assume that the true size of the S-I-R compartments in Pool 2 is not known. To simplify the presentation, we assume that $I_t^{(1)}$ *is observed* in Pool 1, perhaps due to better epidemiological surveillance in that pool.

In our detection problem, the main event of interest is the presence of infecteds above the threshold level in Pool 2, $\{I_t^{(2)} > \bar{I}\}$. Accordingly, we consider $\tilde{P}_t = P(I_t^{(2)} > \bar{I}|\mathcal{G}_t)$, the posterior probability that the epidemic started in Pool 2 at time t given the limited knowledge about it available by time t, here summarized by some information set $\mathcal{G}_t$. Depending on assumptions about the observations structure, $\tilde{P}_t$ may be available in closed form (e.g. through Bayesian conjugate updating [Ludkovski and Niemi, 2010, Lin and Ludkovski, 2014]) or may have to be only approximately computed through e.g. particle filtering methods (see Appendix A) [Sheinson et al., 2014, Skvortsov and Ristic, 2012].

The latter method, which computes the whole posterior distribution $\pi_t \sim I_t^{(2)}|\mathcal{G}_t$, is computationally expensive, while conjugate updating requires carrying several sufficient statistics about the posterior of $I_t^{(2)}$. In either case, $\tilde{P}_t$ on its own is not Markovian, and hence does not possess simple dynamics. Therefore we propose a model that works with a simplified, Markovian version of $\tilde{P}_t$, which we denote as P_t, as well as a model based on Bayesian conjugate updating of the posterior distribution of $I_t^{(2)}$.

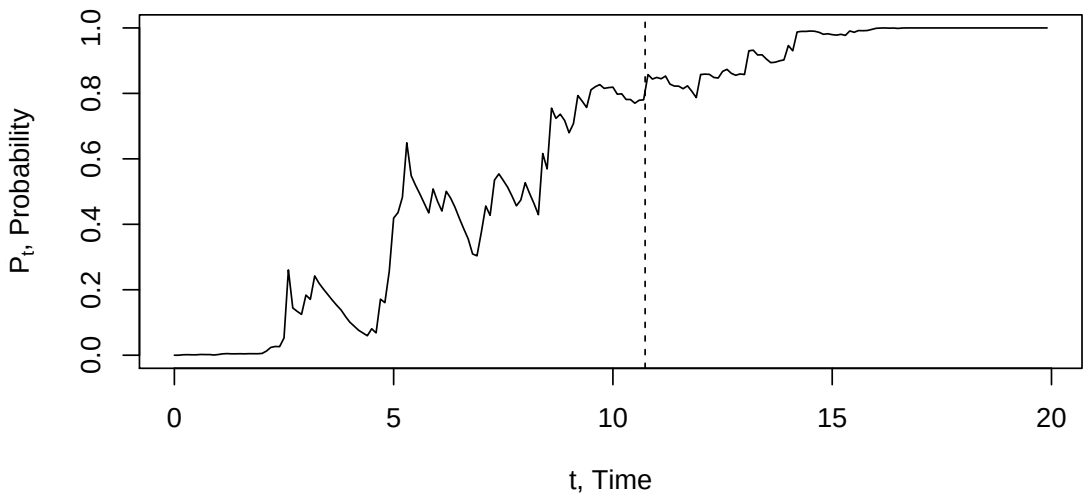


Figure 4.1: Posterior probability that the epidemic started in the second Pool $\tilde{P}_t$ vs. time t, where the dashed line represents the actual start time θ of outbreak in Pool 2. The plot was constructed using particle filtering using the following model parameter values: $\beta = 0.75,\ \alpha = 0.01,\ \gamma = 0.5,\ M^{(1)} = M^{(2)} = 2000$.

Figure 4.1 shows a sample scenario of the evolution of $\tilde{P}_t$ in a partially observed framework. The plot was generated using particle filtering (see Appendix A) and used the two-pool SIR model (that is a combination of reaction channels (3.1) with $\mu_{EI} = \infty$ and reaction channels (3.28)) with noisy Poisson-type observations in each pool [Shatskikh and Ludkovski, 2017]. We observe that $\tilde{P}_t$ tends to drift up (i.e. the posterior probability of outbreak increases over time) and eventually hits 1.

4.1 First Pseudo-Posterior Reduced Model

Our first reduced model consists of the state of epidemic in the first Pool $\left\{S_t^{(1)}, I_t^{(1)}\right\}$ and a process P_t that is interpreted as the probability that the epidemic reached Pool 2 conditional on the information $\mathcal{G}_t = \sigma(S_{0:t}^{(1)}, I_{0:t}^{(1)})$ from Pool 1. The first two components $S_t^{(1)}$ and $I_t^{(1)}$ come from a one-population SIR model (see the definition of the SIR model in Section 3.1.1). To prescribe the dynamics of the pseudo-posterior P_t, we decompose the event $\{I_t^{(2)} > \bar{I}\} \equiv \{\theta \leq t\}$ into two cases: the event that the epidemic already started at time $t-1$ (i.e. $\theta \leq t-1$), and the event that it starts at $t = \theta$. Note, for this particular section we will assume that our epidemic threshold $\bar{I} = 0$, i.e. the epidemic is defined if the number of infected greater than 0. We also add some stochastic noise to denote exogenous fluctuations in our posterior estimates regarding the second pool. In total, we thus assume that

$$P_t = P_{t-1} + P(I_{t-1}^{(2)} = 0 \text{ and } I_t^{(2)} > 0 | \mathcal{G}_{t-1}) + \delta_t, \tag{4.1}$$

where δ_t are i.i.d. noise terms. Intuitively, the probability of outbreak has a positive drift over time, and the drift is precisely the posterior probability of the outbreak beginning during the current period, $\{\theta \in [t-1, t]\}$.

From the SIR dynamics, the probability that $\{\theta \in [t-1, t]\}$ conditional on Pool 1 observations up to previous stage $t-1$ is equal to the product of the probability that an infected from Pool 1 interacts with a susceptible from Pool 2 and the conditional probability that $\{\theta > t-1\}$. The former happens with rate $\mu_{SI}^{(1,2)}(t) I_s^{(1)} \frac{S_s^{(2)}}{M^{(2)}}$, $s \in [t-1, t]$, while the latter event is the complement of $\{\theta \leq t-1\}$ and hence has probability $1 - P_{t-1}$. Using the fact that conditional on $\{\theta \geq t\}$, $M^{(2)} = S_{t-1}^{(2)}$, and making the transition rate

constant on $[t-1,t]$ we obtain

$$P(\theta \in [t-1,t] | \mathcal{G}_{t-1}) \simeq \mu_{SI}^{(1,2)}(t) I_{t-1}^{(1)}(1-P_{t-1}). \tag{4.2}$$

To guarantee $P_t \in [0,1]$ is interpretable as a probability we confine it to $[0,1]$, yielding

$$P_t := \begin{cases} 0 \vee (P_{t-1} + \mu_{SI}^{(1,2)}(t) I_{t-1}^{(1)}(1-P_{t-1}) + \delta_t) \wedge 1, & \text{if } P_{t-1} \neq 1, \\ 1, & \text{if } P_{t-1} = 1. \end{cases} \tag{4.3}$$

In our simulations we use centered Gaussian noise $\delta_t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma_\delta^2)$ with variance σ_δ^2, however it can take any distribution. The algorithm of generating P_t is given in Algorithm B.1.

Note that $P_t = 1$ is an absorbing state, representing certainty that the outbreak reached Pool 2, while $P_t = 0$ is a boundary case since even if it is certain that the outbreak is currently not in Pool 2, it can still get cross-infected in the future. Similar features hold for the true posterior probability $\tilde{P}_t$, cf. Figure 4.1. An alternative model for the probability of outbreak P_t is discussed in Section 4.2.

Remark 3 *Note that* (4.3) *is in discrete-time; to connect to the continuous-time dynamics of SIR one could take the limit as the time increment goes to zero, obtaining a diffusive model* $dP_t = \mu_{SI}^{(1,2)}(t) I_t^{(1)}(1-P_t)\,dt + \delta dW_t$ *where* (W_t) *is a Brownian motion. However, since detection is assumed to take place only at instances* $t = 1, 2, \ldots$*, we prefer to work with* (4.3) *as is.*

4.1.1 Detection Within the First Reduced Model

To sum up, the developed reduced 2-pool model has a 3-dimensional state $\{\mathfrak{X}^{\mathrm{I}}\}_t = \left(S_t^{(1)}, I_t^{(1)}, P_t\right)$ with state space

$$\mathcal{X} := \{(s, i, p) : s, i \in \mathbb{N}, s + i < M^{(1)}, p \in [0, 1]\}.$$

Figure 4.2 shows a few sample trajectories of $\mathfrak{X}^{\mathrm{I}}$ to illustrate the resulting dynamics.

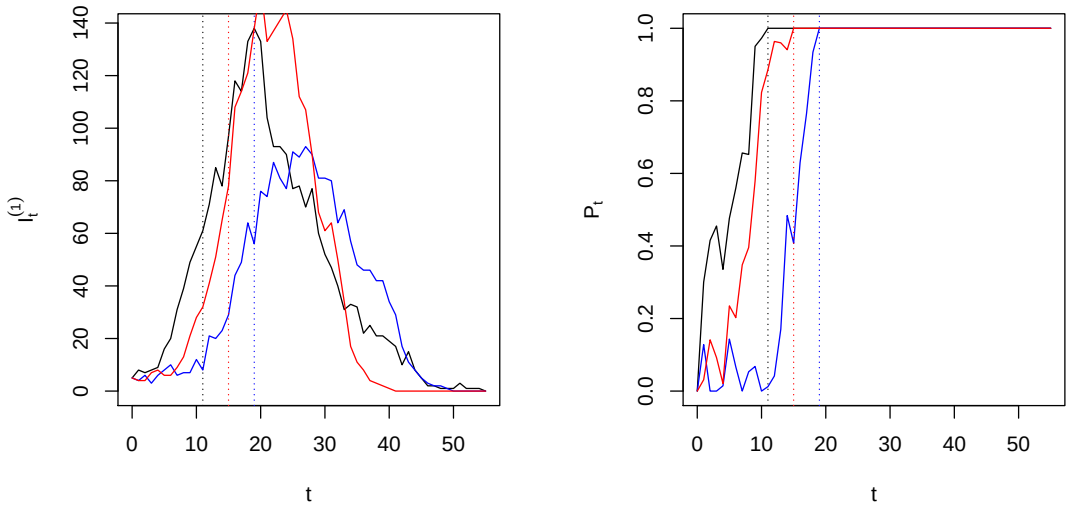


Figure 4.2: Three sample trajectories of $\mathfrak{X}^{\mathrm{I}}$ with the initial condition $S_0^{(1)} = 1995$, $I_0^{(1)} = 5$, $P_0 = 0$ and outbreak parameters from Table 6.1. The left panel is the plot of $\{I_t^{(1)}\}$, the number of infecteds in the first population, and the right panel is the plot of $\{P_t\}$, the posterior probability that the epidemic started in the second population. The vertical dotted lines represent times when P_t hits 1 and outbreak becomes certain.

Our detection problem (2.10) relies on the computation of the immediate and future expected costs $E\left[c(\mathfrak{X}^{\mathrm{I}}_{0:\tau})|\mathfrak{X}^{\mathrm{I}}_0\right]$ and $d(\mathfrak{X}^{\mathrm{I}}_0)$. Re-writing the definitions of immediate and future costs (2.2) and (2.3) in terms of the event $\{I_t^{(2)} > 0\}$, and taking conditional

expectation we obtain:

$$d(\mathfrak{X}_0^{\mathrm{I}}) := C_{\mathrm{FA}}(1 - P_0), \tag{4.4}$$

$$c(\mathfrak{X}_{0:\tau}^{\mathrm{I}}) := \sum_{s=0}^{\tau-1} C_{\mathrm{Delay}} P_s + C_{\mathrm{FA}}(1 - P_\tau), \tag{4.5}$$

where $\tau \in \mathcal{S}$. Rather than in terms of the unobserved $I^{(2)}$, the above expressions are now given in terms of the component P_t, allowing us to measure detection costs within the $\mathfrak{X}^{\mathrm{I}}$-model. Notice that $d(\mathfrak{X}_0^{\mathrm{I}})$ is a function of P_0 and $c(\mathfrak{X}_{0:\tau}^{\mathrm{I}})$ is a function of the future trajectory $\{P_s : s = 0, \ldots, \tau\}$.

4.2 Second Pseudo-Posterior Reduced Model

We assume that we observe the state of epidemic in Pool 1

$$\mathbf{X}_{0:t}^{(1)} := \left\{ S_{0:t}^{(1)}, E_{0:t}^{(1)}, I_{0:t}^{(1)}, R_{0:t}^{(1)} \right\},$$

the number of observed cases in Pool 2 $C_{0:t}^{obs(2)}$ as well as all parameters of Two Population Asymmetrical SEIR model. The number of Infectious individuals in Pool 2 is computed via:

$$I_t^{(2)} = I_{t-1}^{(2)} + \Delta I_t^{(2)} - \Delta R_t^{(2)} - \Delta(DI)_t^{(2)}, \tag{4.6}$$

where $\Delta I_t^{(2)}$ are the new infections, $\Delta R_t^{(2)}$ are the new recovereds, and $\Delta(DI)_t^{(2)}$ are the deceased individuals at discrete time $t = 1, 2, \ldots$. As it was discussed in Chapter 3.1.1, the distribution of $\Delta R_t^{(2)}$ and $\Delta(DI)_t^{(2)}$ is Poisson with respective means $\mu_{IR} I_t^{(2)}$ and $\mu_D I_t^{(2)}$, where μ_{IR} is the recovery rate.

If we assume that all infections in the second populattion are self-infections (i.e. the susceptibles were infected by the infected in the same Pool), then the new infecteds in the

second Pool at time t consist of cross-infections from the first Pool $\Delta I_t^{(1\to 2)}$, observable self-infections in the second Pool $C_t^{obs(2)}$, and non-observable self-infections in the second Pool $C_t^{non-obs(2)}$

$$\Delta I_t^{(2)} = \Delta I_t^{(1\to 2)} + C_t^{obs(2)} + C_t^{non-obs(2)}, \tag{4.7}$$

where cross-infections and self-infections are respectively distributed as

$$\Delta I_t^{(1\to 2)} \sim Poisson(\mu_{SI}^{(1,2)}(t)\eta_t^{(2)} I_t^{(1)}),$$

$$C_t^{obs(2)} \sim Poisson((1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)} I_t^{(2)}), \tag{4.8}$$

which was also defined in (3.14), and

$$C_t^{non-obs(2)} \sim Poisson(\rho\mu_{SI}^{(2)}(t)\eta_t^{(2)} I_t^{(2)}), \tag{4.9}$$

where we defined

$$\eta_t^{(2)} = \frac{S_t^{(2)}}{M_t^{(2)}}.$$

Remark 4 *An alternative decomposition of $\Delta I_t^{(2)}$ is given in the Appendix A, where we propose that the change in the Infected individuals consists of observed cases and noisy observations (false positive cases).*

Thus, combining equations (4.6) and (4.7) we get:

$$I_t^{(2)} = I_{t-1}^{(2)} + \Delta I_t^{(1\to 2)} + C_t^{non-obs(2)} + C_t^{obs(2)} - \Delta R_t^{(2)}, \tag{4.10}$$

Therefore, we can compute expected number of infected individuals in Pool 2 at time t given the number of infecteds individuals in Pool 1 at time t and in Pool 2 at time $t-1$

assuming independence of $\eta_t^{(2)}$ and $I_t^{(1)}$ as well as independence of $\eta_t^{(2)}$ and $I_t^{(2)}$:

$$E(I_t^{(2)}|I_t^{(1)}, I_{t-1}^{(2)}) = I_{t-1}^{(2)} + \mu_{SI}^{(1,2)}(t)\eta_t^{(2)} I_t^{(1)} + \rho\mu_{SI}^{(2)}(t)\eta_t^{(2)} I_t^{(2)} + (1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)} I_t^{(2)} - \mu_{IR} I_t^{(2)} \tag{4.11}$$

Note that the independence of $\eta_t^{(2)}$ and $I_t^{(2)}$ is satisfied when the number of Susceptible individuals is approximately equal to the size of population individuals, or $\eta_t^{(2)} \approx 1$. However in our calculations we treat $\eta_t^{(2)}$ as known.

Theorem 1 *If the likelihood*

$$C_{t+1}^{obs(2)}|\mathbf{X}_{0:t}^{(1)}, I_t^{(2)}, C_{0:t}^{obs(2)} \sim Poisson\left((1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)} I_t^{(2)}\right)$$

and the prior

$$I_t^{(2)}|\mathbf{X}_{0:t}^{(1)}, C_{0:t}^{obs(2)} \sim Gamma(\alpha_t, \beta_t),$$

then

$$I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)} \sim Gamma\left(\alpha_t + C_{t+1}^{obs(2)}, \beta_t + (1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}\right).$$

Proof: The pdf of $C_{t+1}^{obs(2)}|\mathbf{X}_{0:t}^{(1)}, I_t^{(2)}, C_{0:t}^{obs(2)}$ is defined as

$$p_{C_{t+1}^{obs(2)}|\mathbf{X}_{0:t+1}^{(1)}, I_t^{(2)}, C_{0:t}^{obs(2)}}\left(c_{t+1}|\mathbf{x}_{0:t+1}^{(1)}, i_t, c_{0:t}\right)$$

and it is equal to

$$\frac{\left((1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)} i_t\right)^{c_{t+1}}}{c_{t+1}!} e^{-(1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)} i_t}, \quad c_t \in \mathcal{N}. \tag{4.12}$$

The pdf of $I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t}^{obs(2)}$ is equal to the pdf of $I_t^{(2)}|\mathbf{X}_{0:t}^{(1)}, C_{0:t}^{obs(2)}$ and it is equal to

$$\pi_{I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t}^{obs(2)}}\left(i_t|\mathbf{x}_{0:t+1}^{(1)}, c_{0:t}\right) = \frac{\beta_t^{\alpha_t}}{\Gamma(\alpha_t)} i_t^{\alpha_t - 1} e^{-\beta_t i_t}, \quad i_t \in \mathcal{R}^+. \tag{4.13}$$

We can compute the posterior distribution of $I_t^{(2)}$ using Bayes rule:

$$p_{I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)},C_{0:t+1}^{obs(2)}}\left(i_t|\mathbf{x}_{0:t+1}^{(1)},c_{0:t+1}^{(2)}\right) = \frac{p_{C_{t+1}^{obs(2)}|\mathbf{X}_{0:t+1}^{(1)},I_t^{(2)},C_{0:t}^{obs(2)}}\left(c_{t+1}|\mathbf{x}_{0:t+1}^{(1)},i_t,c_{0:t}\right)\pi_{I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)},C_{0:t}^{obs(2)}}\left(i_t|\mathbf{x}_{0:t+1}^{(1)},c_{0:t}\right)}{\int p_{C_{t+1}^{obs(2)}|\mathbf{X}_{0:t+1}^{(1)},I_t^{(2)},C_{0:t}^{obs(2)}}\left(c_{t+1}|\mathbf{x}_{0:t+1}^{(1)},i_t,c_{0:t}\right)\pi_{I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)},C_{0:t}^{obs(2)}}\left(i_t|\mathbf{x}_{0:t+1}^{(1)},c_{0:t}\right)\mathbf{d}i_t}. \quad (4.14)$$

First, the numerator of equation (4.14) is just a product of densities (4.12) and (4.13):

$$\frac{\beta_t^{\alpha_t}\left((1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}\right)^{c_{t+1}}}{\Gamma(\alpha_t)c_{t+1}!}i_t^{c_{t+1}+\alpha_t-1}e^{-\left(\beta_t+(1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}\right)i_t} \quad c_t\in\mathcal{N}, i_t\in\mathcal{R}^+. \quad (4.15)$$

Second, the denominator of equation (4.14) is just an integral of Equation (4.15) with respect to $I_t^{(2)}$:

$$\begin{aligned}
&\int_0^\infty \frac{\beta_t^{\alpha_t}\left((1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}\right)^{c_{t+1}}}{\Gamma(\alpha_t)c_{t+1}!}i_t^{c_{t+1}+\alpha_t-1}e^{-\left(\beta_t+(1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}\right)i_t}\mathbf{d}i_t \\
&= \frac{\beta_t^{\alpha_t}\left((1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}\right)^{c_{t+1}}}{\Gamma(\alpha_t)c_{t+1}!}\int_0^\infty i_t^{c_{t+1}+\alpha_t-1}e^{-\left(\beta_t+(1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}\right)i_t}\mathbf{d}i_t \\
&= \frac{\beta_t^{\alpha_t}\left((1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}\right)^{c_{t+1}}}{\Gamma(\alpha_t)c_{t+1}!}\frac{\Gamma(c_{t+1}+\alpha_t)}{\left(\beta_t+(1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}\right)^{c_{t+1}+\alpha_t}}, \quad c_t\in\mathcal{N}, \quad (4.16)
\end{aligned}$$

where to go from the second to last line using the fact that pdf of Gamma distributed random variable is equal to 1.

We computed the numerator in the Equation (4.15) and the denominator in the Equation (4.16), thus we finally get the posterior density of $I_t^{(2)}$ to be

$$\frac{\left(\beta_t+(1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}\right)^{c_{t+1}+\alpha_t}}{\Gamma(c_{t+1}+\alpha_t)}i_t^{c_{t+1}+\alpha_t-1}\exp\left\{-\left(\beta_t+(1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}\right)i_t\right\}, \quad (4.17)$$

which is the density of Gamma distributed random variable with parameters $\alpha_t + c_{t+1}$ and $\beta_t + (1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}$. ■

Theorem 2 *If we define*

$$\mu_{t+1} := E(I_{t+1}^{(2)} \,|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)})$$

and

$$\sigma_{t+1}^2 = Var\left(I_{t+1}^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right)$$

and using the assumptions of Theorem 1, we can compute these quantities as:

$$\begin{aligned}
\mu_{t+1} &= (1+\rho\mu_{SI}^{(2)}(t)\eta_t^{(2)} - \gamma - \mu_D)\frac{\alpha_t + C_{t+1}^{obs(2)}}{\beta_t + (1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}} + \mu_{SI}^{(1,2)}(t)\eta_t^{(2)} I_t^{(1)} + C_{t+1}^{obs(2)}, \\
\sigma_{t+1}^2 &= \frac{(\alpha_t + C_{t+1}^{obs(2)})(1+\gamma^2+\mu_D^2+\rho^2(\mu_{SI}^{(2)}(t))^2(\eta_t^{(2)})^2)}{\left(\beta_t + (1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}\right)^2} \\
&\quad + \frac{(\alpha_t + C_{t+1}^{obs(2)})(\gamma + \mu_D + \rho\mu_{SI}^{(2)}(t)\eta_t^{(2)})}{\beta_t + (1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}} + \mu_{SI}^{(1,2)}(t)\eta_t^{(2)} I_t^{(1)}.
\end{aligned}$$

Proof: We start by computing the posterior mean μ_{t+1}. First, by definition of $I_t^{(2)}$ (see Equation (4.10)) we get:

$$\mu_{t+1} = E\left(I_t^{(2)} - \Delta R_t^{(2)} - \Delta(DI)_t^{(2)} + \Delta I_{t+1}^{(1\to 2)} + C_{t+1}^{non-obs(2)} + C_{t+1}^{obs(2)} \mid \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right).$$

Then using the linearity property of expectation we get:

$$\begin{aligned}
\mu_{t+1} = E\left(I_t^{(2)} \mid \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) - E\left(\Delta R_t^{(2)} + \Delta(DI)_t^{(2)} \mid \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) \\
+ E\left(\Delta I_{t+1}^{(1\to 2)} + C_{t+1}^{non-obs(2)} + C_{t+1}^{obs(2)} \mid \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right).
\end{aligned}$$

Using the tower property we get

$$\mu_{t+1} = E\left(I_t^{(2)} \mid \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) - (\gamma+\mu_D)E\left(I_t^{(2)} \mid \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) + \mu_{SI}^{(1,2)}(t)\eta_t^{(2)} I_t^{(1)} + \rho\mu_{SI}^{(2)}(t)\eta_t^{(2)} E\left(I_t^{(2)} \mid \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) + C_{t+1}^{obs(2)}.$$

Thus, we can rewrite it as:

$$\mu_{t+1} = (1+\rho\mu_{SI}^{(2)}(t)\eta_t^{(2)} - \gamma - \mu_D)E\left(I_t^{(2)} \mid \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) + \mu_{SI}^{(1,2)}(t)\eta_t^{(2)} I_t^{(1)} + C_{t+1}^{obs(2)}. \quad (4.18)$$

While computing the posterior variance one needs to be more careful because the variance of the sum is equal to sum of variance only if uncorrelated random variables are to be summed up. However it might be observed that all our random variables in our computation are uncorrelated and therefore we don't have correlation terms.

First, we use the definition of $I_t^{(2)}$ (see Equation (4.10)) to get:

$$\sigma_{t+1}^2 = Var\left(I_t^{(2)} - \Delta R_t^{(2)} + \Delta I_t^{(1\to 2)} + C_t^{non-obs(2)} + C_t^{obs(2)} - \Delta(DI)_t^{(2)} | \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right).$$

Second, we get:

$$\begin{aligned}\sigma_{t+1}^2 = {} & Var\left(I_t^{(2)} | \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) + Var\left(\Delta R_t^{(2)} | \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) \\ & + Var\left(\Delta I_t^{(1\to 2)} | \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) + Var\left(C_t^{non-obs(2)} | \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) \\ & + Var\left(C_t^{obs(2)} | \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) + Var\left(\Delta(DI)_t^{(2)} | \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right), \end{aligned} \quad (4.19)$$

where

$$Var\left(\Delta I_t^{(1\to 2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) = \mu_{SI}^{(1,2)}(t)\eta_t^{(2)} I_t^{(1)}, \tag{4.20}$$

$$Var\left(C_t^{obs(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) = 0. \tag{4.21}$$

To compute the other three variances $Var\left(C_t^{non-obs(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right)$, $Var\left(\Delta(DI)_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right)$, and $Var\left(\Delta R_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right)$ we need to use the law of total variance as well as other basic properties of expectation and variance:

$$\begin{aligned}
&Var\Big(C_t^{non-obs(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\Big) \\
&\quad = E\left(Var\left(C_t^{non-obs(2)}|I_t^{(2)}, \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right)|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) \\
&\qquad + Var\left(E\left(C_t^{non-obs(2)}|I_t^{(2)}, \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right)|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) \\
&\quad = E\left(\rho\mu_{SI}^{(2)}(t)\eta_t^{(2)} I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) + Var\left(\rho\mu_{SI}^{(2)}(t)\eta_t^{(2)} I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) \\
&\quad = \rho\mu_{SI}^{(2)}(t)\eta_t^{(2)} E\left(I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) + \rho^2\mu_{SI}^{(2)}(t)^2(\eta_t^{(2)})^2 Var\left(I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right),
\end{aligned} \tag{4.22}$$

$$\begin{aligned}
&Var\Big(\Delta(DI)_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\Big) \\
&\quad = E\left(Var\left(\Delta(DI)_t^{(2)}|I_t^{(2)}, \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right)|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) \\
&\qquad + Var\left(E\left(\Delta(DI)_t^{(2)}|I_t^{(2)}, \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right)|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) \\
&\quad = E\left(\mu_D I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) + Var\left(\mu_D I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) \\
&\quad = \mu_D E\left(I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) + \mu_D^2 Var\left(I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right),
\end{aligned} \tag{4.23}$$

and

$$
\begin{aligned}
&Var\left(\Delta R_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) \\
&\quad = E\left(Var\left(\Delta R_t^{(2)}|I_t^{(2)}, \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right)|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) \\
&\qquad + Var\left(E\left(\Delta R_t^{(2)}|I_t^{(2)}, \mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right)|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) \\
&\quad = E\left(\gamma I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) + Var\left(\gamma I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) \\
&\quad = \gamma E\left(I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) + \gamma^2 Var\left(I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right). \qquad (4.24)
\end{aligned}
$$

Therefore, inserting the results from Equations (4.20), (4.21), (4.22), (4.23), and (4.24) into Equation (4.19) we get

$$
\begin{aligned}
\sigma_{t+1}^2 &= Var\left(I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) + (\gamma + \mu_D)E\left(I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) \\
&\quad + (\gamma^2 + \mu_D^2)Var\left(I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) + \mu_{SI}^{(1,2)}(t)\eta_t^{(2)} I_t^{(1)} \\
&\quad + \rho\mu_{SI}^{(2)}(t)\eta_t^{(2)} E\left(I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) + \rho^2\mu_{SI}^{(2)}(t)^2(\eta_t^{(2)})^2 Var\left(I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) \\
&= (1 + \gamma^2 + \mu_D^2 + \rho^2\mu_{SI}^{(2)}(t)^2(\eta_t^{(2)})^2)Var\left(I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) \\
&\quad + (\gamma + \mu_D + \rho\mu_{SI}^{(2)}(t)\eta_t^{(2)})E\left(I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) + \mu_{SI}^{(1,2)}(t)\eta_t^{(2)} I_t^{(1)}. \qquad (4.25)
\end{aligned}
$$

By Theorem 1 $I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)} \sim Gamma(\alpha_t + c_{t+1}, \beta_t + (1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)})$ and the mean and variance of Gamma random variable we get:

$$
E\left(I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) = \frac{\alpha_t + C_{t+1}^{obs(2)}}{\beta_t + (1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}}, \qquad (4.26)
$$

$$
Var\left(I_t^{(2)}|\mathbf{X}_{0:t+1}^{(1)}, C_{0:t+1}^{obs(2)}\right) = \frac{\alpha_t + C_{t+1}^{obs(2)}}{\left(\beta_t + (1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}\right)^2}. \qquad (4.27)
$$

Substituting the results (4.26) and (4.27) into equations (4.18) and (4.25) we get

$$\mu_{t+1} = (1 + \rho\mu_{SI}^{(2)}(t)\eta_t^{(2)} - \gamma - \mu_D)\frac{\alpha_t + C_{t+1}^{obs(2)}}{\beta_t + (1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}} + \mu_{SI}^{(1,2)}(t)\eta_t^{(2)} I_t^{(1)} + C_{t+1}^{obs(2)} \tag{4.28}$$

$$\begin{aligned}\sigma_{t+1}^2 = &\frac{(\alpha_t + C_{t+1}^{obs(2)})(1 + \gamma^2 + \mu_D^2 + \rho^2(\mu_{SI}^{(2)}(t))^2(\eta_t^{(2)})^2)}{\left(\beta_t + (1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}\right)^2} \\ &+ \frac{(\alpha_t + C_{t+1}^{obs(2)})(\gamma + \mu_D + \rho\mu_{SI}^{(2)}(t)\eta_t^{(2)})}{\beta_t + (1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}} + \mu_{SI}^{(1,2)}(t)\eta_t^{(2)} I_t^{(1)}.\end{aligned} \tag{4.29}$$

■

Theorem 2 gives us the updates of μ_{t+1} and σ_{t+1}^2. We can solve for α_t and β_t from μ_t and σ_t^2:

$$\alpha_t = \frac{\mu_t^2}{\sigma_t^2}, \qquad \beta_t = \frac{\mu_t}{\sigma_t^2}. \tag{4.30}$$

Therefore with some initial conditions μ_0, σ_0^2, $\mathbf{X}_{0:T}^{(1)}$, $C_{0:T}^{obs(2)}$, where T is the end time, we can generate the process $\{\mu_t, \sigma_t^2\}, t = 0, \ldots, T$ using Algorithm B.2, which we denote as 'Original' procedure.

Figure 4.3 shows the Infected Individuals in Bristol, which is the original observed data of weekly cases in Bristol from Section 3.3 multiplied by $1/\rho$ to get the approximated number of Infected individuals, and 95% quantile region of the Infected Individuals $I_t^{(2)}$, which was generated using the starting conditions $\mu_0 = 200$ and $\sigma_0^2 = 100$ via Algorithm B.2 . We assumed that the Bristol city is isolated from others and therefore there are no cross-infections ($\mu_{SI}^{(1,2)}(t) = 0$), and therefore we didn't use $\mathbf{X}_{0:T}^{(1)}$. We observe that only 60% of points are within the 95% Quantile Intervals, however both points and quantile intervals roughly follow the same curve.

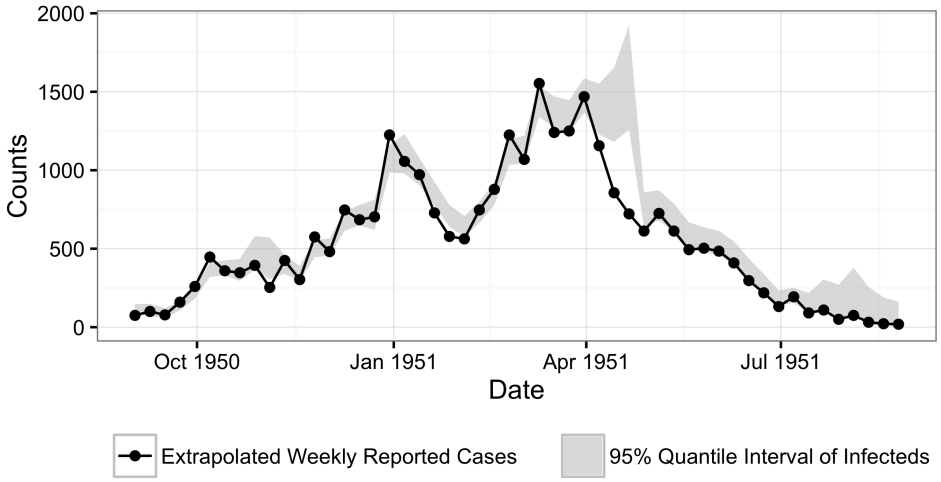


Figure 4.3: Time Series of Extrapolated Measles Epidemic in Bristol in 1950-1951 (solid line) in comparison to the 95% predicted quantile region (shaded region) of the Infected Individuals $I_t^{(2)}$ using the Second Reduced Model with $\mu_{SI}^{(1,2)}(t) = 0$

The values of $C^{obs(2)}$ might not be available. So we propose a procedure of generating $\{\mu_t, \sigma_t^2\}, t = 0, 1, 2 \ldots, T$ without using the observed cases in the second population $C_{0:T}^{obs(2)}$. Instead we generate the value of $I_t^{(2)}$ from a posterior Gamma distribution with parameters given in Equation (4.30) and sample $C_t^{obs(2)}$ from a Poisson distribution with rate (4.8). The algorithm of this procedure is given in Algorithm (B.3), and we refer to this Algorithm as 'Gamma-Poisson' algorithm.

Theorem 3 *If*

$$I_t^{(2)} | \mathbf{X}_{0:t}^{(1)}, C_{0:t}^{obs(2)} \sim Gamma(\alpha_t, \beta_t)$$

and

$$C_t^{obs(2)} \sim Poisson((1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)} I_t^{(2)}),$$

then $C_t^{obs(2)} | \mathbf{X}_{0:t-1}^{(1)}, I_t^{(2)}, C_{0:t-1}^{obs(2)}$ *has a Negative Binomial distribution with parameters* α_t *and* $\frac{\beta_t}{(1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)} + \beta_t}$.

Proof: The density of

$$f_{C_t^{obs(2)}|\mathbf{X}_{0:t}^{(1)},C_{0:t-1}^{obs(2)}}(c) = \int_0^\infty f_{C_t^{obs(2)},I_t^{(2)}|\mathbf{X}_{0:t}^{(1)},C_{0:t-1}^{obs(2)}}(c,i)\mathrm{d}i$$
$$= \int_0^\infty f_{C_t^{obs(2)}|\mathbf{X}_{0:t-1}^{(1)},I_t^{(2)},C_{0:t-1}^{obs(2)}}(c) f_{I_t^{(2)}|\mathbf{X}_{0:t}^{(1)},C_{0:t-1}^{obs(2)}}(i)\mathrm{d}i. \tag{4.31}$$

The probability mass function of $C_t^{obs(2)}|\mathbf{X}_{0:t-1}^{(1)}, I_t^{(2)}, C_{0:t-1}^{obs(2)}$ is given in Equation (4.12). Substituting this density and the density of $Gamma(\alpha_t, \beta_t)$ into Equation (4.31) we get

$$f_{C_t^{obs(2)}|\mathbf{X}_{0:t}^{(1)},C_{0:t-1}^{obs(2)}}(c) \int_0^\infty e^{-(1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}i)} \frac{\left((1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}i\right)^c}{c!} \frac{\beta_t^{\alpha_t}}{\Gamma(\alpha_t)} i_t^{\alpha_t-1} e^{-\beta_t i_t}\mathrm{d}i$$
$$= \left((1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}\right)^c \frac{\beta_t^{\alpha_t}}{(\alpha_t-1)!\,c!} \int_0^\infty i^{c+\alpha_t-1} e^{-(\beta_t+(1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)})i}\mathrm{d}i.$$

Using the fact that

$$\int_0^\infty \frac{(\beta_t+(1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)})^{c+\alpha_t}}{\Gamma(c+\alpha_t)} i^{c+\alpha_t-1} e^{-(\beta_t+(1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)})i}\mathrm{d}i = 1$$

we get the density $f_{C_t^{obs(2)}|\mathbf{X}_{0:t}^{(1)},C_{0:t-1}^{obs(2)}}(c)$ equal to

$$\frac{\Gamma(c+\alpha_t)}{(\alpha_t-1)!\,c!} \left((1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}\right)^c \beta_t^{\alpha_t} (\beta_t+(1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)})^{-c-\alpha_t}. \tag{4.32}$$

Rearranging the terms in Equation (4.32) we get:

$$\frac{(c+\alpha_t-1)!}{(\alpha_t-1)!\,c!} \left(\frac{\beta_t}{\beta_t+(1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}}\right)^{\alpha_t} \left(1-\frac{\beta_t}{\beta_t+(1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}}\right)^c, \tag{4.33}$$

which is a probaility mass function of a Negative Binomial distribution with parameters α_t and $\frac{\beta_t}{(1-\rho)\mu_{SI}^{(2)}(t)\eta_t^{(2)}+\beta_t}$. ■

Theorem 3 helps us to avoid a generation of two values $I_t^{(2)}$ and $C_t^{(2)}$. We only sample

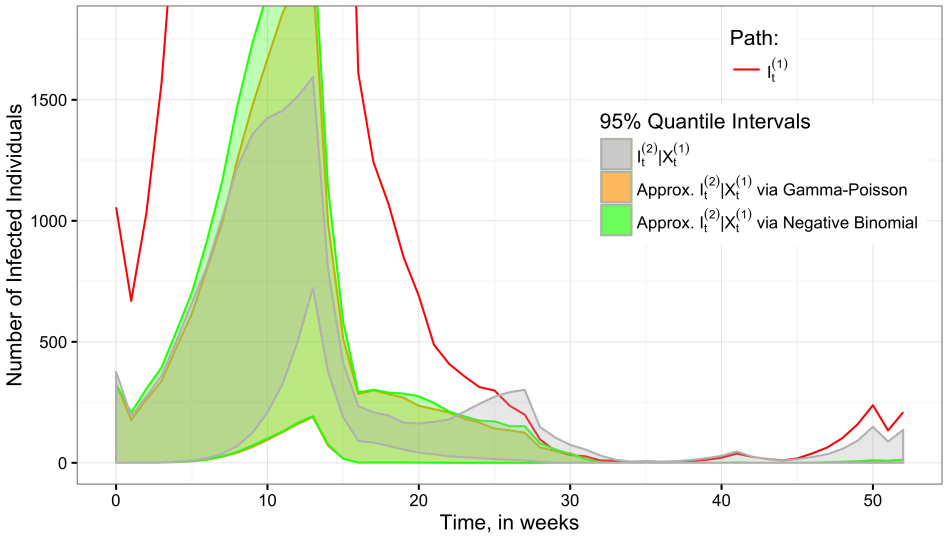


Figure 4.4: Comparison of the true $I_t^{(2)}|X_t^{(1)}$ to approximated $I_t^{(2)}|X_t^{(1)}$ computed using the 'Gamma-Poisson', and 'Negative Binomial' algorithms. The quantile intervals were computed based on 1000 trajectories from the true $I_t^{(2)}|X_t^{(1)}$ as well as approximations of $I_t^{(2)}|X_t^{(1)}$.

$C_t^{(2)}$ from a Negative Binomial distribution. Our procedure simplifies to Algorithm (B.4), and we refer to this Algorithm as 'Negative Binomial'. So both Algorithms 'Gamma-Poisson' and 'Negative Binomial' should give the same results, and we give a preferenace to 'Negative Binomial' from a theoretical point of view.

We generate a path $\{\mathbf{X}_t^{(1)}, \mathbf{X}_t^{(2)}\}$ for a year (52 weeks) from a Two Population Asymmetrical SIR model (see Section 3.2). Given the path of $\mathbf{X}_t^{(1)}$ as well as the starting points $\mu_0 = 50$ and $\sigma_0 = 100$ we generate the paths $\{\mu_t, \sigma_t\}$ from the 'Gamma-Poisson', and 'Negative Binomial' algorithms, then solve for $\{\alpha_t, \beta_t\}$ and use these parameters to sample $I_t^{(2)}$ from a Gamma distribution. We also compute the possible trajectories of I_t^2 given $X_t^{(1)}$, where as a starting position of the epidemic in the second population we use a sample from Gamma posterior with $\mu_0 = 50$ and $\sigma_0^2 = 100$.

Figure 4.4 compares the 95% quantile intervals of true $I_t^{(2)}|X_t^{(1)}$ to the 95% quantile intervals of $I_t^{(2)}|X_t^{(1)}$ approximations computed using the 'Gamma-Poisson' and 'Negative

Binomial' algorithms. First, notice that the path $I_t^{(1)}$ has two peaks: the first one is a larger one reaching level of 1500 infecteds and the second is a smaller one reaching level of 300. Second, 95% Quantile Regions computed using the 'Gamma-Poisson' and 'Negative Binomial' algorithms are almost the same. That is expected by the construction of the Algorithms. Third, the 95% quantile intervals of the true $I_t^{(2)}|X_t^{(1)}$ from week 0 to week 22 are fully inside both 95% quantile intervals of approximations, however neither the 'Gamma-Poisson' nor the 'Negative Binomial' algorithms depicted the second smaller peak of infecteds because they only "see" the values of $I_t^{(1)}$, which didn't rise at that time. Thus, we continue with the Negative Binomial' algorithm for generating a path of $\{\mu_t, \sigma_t\}$.

Finally we define our Second Reduced Model, which consists of the state of epidemic in the first Pool $\{S_t^{(1)}, E_t^{(1)}, I_t^{(1)}, R_t^{(1)}\}$ and a bivariate process $\{\mu_t, \sigma_t^2\}$, that are parameters of gamma posterior distribution of $I_t^{(2)}$, the number of infecteds in the second Pool.

4.2.1 Detection Within the Second Reduced Model

The Second Reduced Model has a 6-dimensional state $\{\mathfrak{X}^{\text{II}}\}_t = \left(S_t^{(1)}, E_t^{(1)}, I_t^{(1)}, R_t^{(1)}, \mu_t, \sigma_t^2\right)$ with state space

$$\mathcal{X} := \{(s, e, i, r, m, v) : s, e, i, r, \in \mathbb{N}, m, v \in \mathbb{R}\}.$$

Our detection problem (2.10) relies on the computation of the immediate and future expected costs $E\left[c(\mathfrak{X}_{0:\tau}^{\text{II}})|\mathfrak{X}_0^{\text{II}}\right]$ and $d(\mathfrak{X}_0^{\text{II}})$. Re-writing the definitions of immediate and future costs (2.2) and (2.3) in terms of the event $\{I_t^{(2)} > \bar{I}\}$, and taking conditional

expectation we obtain:

$$d(\mathfrak{X}_0^{\mathrm{II}}) := C_{\mathrm{FA}}\Psi(\mu_0, \sigma_0^2, \bar{I}), \tag{4.34}$$

$$c(\mathfrak{X}_{0:\tau}^{\mathrm{II}}) := \sum_{s=0}^{\tau-1} C_{\mathrm{Delay}}(1 - \Psi(\mu_s, \sigma_s^2, \bar{I})) + C_{\mathrm{FA}}\Psi(\mu_\tau, \sigma_\tau^2, \bar{I}), \tag{4.35}$$

where $\tau \in \mathcal{S}$ and $\Psi(\mu_t, \sigma_t^2, \bar{I})$ is the cdf of Gamma distribution with the parameters $\{\alpha_t, \beta_t\}$ given in Equation (4.30) evaluated at the point $\bar{I}$. Rather than in terms of the unobserved $I^{(2)}$, the above expressions are now given in terms of the component $\{\mu_t, \sigma_t^2\}$, allowing to measure detection costs within the $\mathfrak{X}^{\mathrm{II}}$-model. Notice that as with the First Reduced Model, $d(\mathfrak{X}_0^{\mathrm{II}})$ is a function of $\{\mu_0, \sigma_0^2\}$ and $c(\mathfrak{X}_{0:\tau}^{\mathrm{II}})$ is a function of the future trajectory $\{\mu_s, \sigma_s^2\}, s = 0, \ldots, \tau$.

Chapter 5

Sequential Regression Monte Carlo

Our goal is to find the detection maps $\hat{\mathfrak{S}}_t$ for $t = 1, 2, \ldots$, defined recursively in (2.10). To do so, at each step, we need to evaluate $E\left[V(t-1, \mathfrak{X}_1)|\mathfrak{X}_0 = \mathbf{x}\right]$ and $d(\mathbf{x})$. The immediate cost $d(\mathbf{x})$ can be computed exactly via (4.4). However, the expectation $E\left[V(t-1, \mathfrak{X}_1)|\mathfrak{X}_0 = \mathbf{x}\right]$ can not be computed analytically since there are no closed-form expressions for the distribution of $\mathfrak{X}_{0:\tau}$. In this Chapter we present the sequential Regression Monte Carlo approach [Egloff, 2005, Gramacy and Ludkovski, 2015] which offers an efficient way to empirically estimate $\hat{\mathfrak{S}}_t$ based on synthetically generated epidemic scenarios. We then use Model Predictive Control to estimate the stationary detection map $\mathfrak{S}$.

5.1 Regression Monte Carlo

For the remainder of this section the auxiliary "time" variable t is fixed and the goal is to approximate the conditional expectation $q(t, \mathbf{x}) := E[c(\mathfrak{X}_{0:\tau^{(t)}}|\mathfrak{X}_0 = \mathbf{x}] - d(\mathbf{x})$ in (2.11). Recall that at step t, detection rules are restricted to satisfy $\tau^{(t)} \leq t$. The Regression Monte Carlo technique approximates $q(t, \cdot)$ by a predicted surrogate value $\hat{q}(t, \cdot)$ which is based on a statistical regression framework.

The surrogate prediction is built using data simulated from the specified model. To do so, a design $\mathcal{Z} := \{\mathbf{x}_0^n,\ n = 1, \ldots, N\}$ of N locations is first generated. Next, we generate the corresponding scenarios $\{\mathfrak{X}_{0:t}^n\}$ with the initial value $\mathfrak{X}_0^n = \mathbf{x}_0^n$, one scenario for each initial location. Define

$$\tau_t^n := \min\{s \geq 1 : \mathfrak{X}_s^n \in \mathfrak{S}_{t-s}\}, \tag{5.1}$$

which leads to the difference of path-wise waiting costs and immediate costs $q^n := c(\mathfrak{X}_{0:\tau_t^n}^n) - d(\mathfrak{X}_0^n)$ using formula (2.3) and (2.2) on the n-th scenario. The aggregate dataset is

$$Z = \{(\mathbf{x}_0^n, q^n), n = 1, \ldots, N\}. \tag{5.2}$$

The construction of $\hat{q}(t, \cdot)$ then involves response surface modeling, i.e. determining the relationship between the initial condition $\mathbf{x}$ and the mean of the sampled $Q|\mathbf{x} \equiv c(\mathfrak{X}_{0:\tau_t}) - d(\mathfrak{X}_0)$. Statistically, we start with

$$Q|\mathbf{x} = q(t, \mathbf{x}) + \epsilon, \tag{5.3}$$

where $q(t, \cdot)$ is the true response surface, $Q = c(\mathfrak{X}_{0:\tau^{(t)}}) - d(\mathfrak{X}_0)$ are random scenario-based difference of costs, and ϵ are mean-zero residuals with variance σ^2 arising from Monte Carlo simulations. Empirically, (5.3) translates into regressing $\{q^n\}$ on $\{\mathbf{x}_0^n\}$, $n = 1, \ldots, N$; this step is discussed in section 5.2. After determining $\hat{q}$, and using (2.10) the estimated detection rule $\hat{\mathfrak{S}}_t$ is

$$\hat{\mathfrak{S}}_t := \{\mathbf{x} : \hat{q}(t, \mathbf{x}) > 0\}. \tag{5.4}$$

The above provides a recipe to obtain an (approximate) $\hat{\mathfrak{S}}_t$ using the collection of

detection rules $\hat{\mathfrak{S}}_{1:t-1}$. Iterating over t yields the sequence of detection maps $\hat{\mathfrak{S}}_t$ for $t = 1, 2, \ldots$. Next, as discussed in Section 2.2 we employ Model Predictive Control (MPC) to achieve the convergence $\mathfrak{S}_t \to \mathfrak{S}$.

5.2 Regression Model

Because we have limited a priori knowledge about the structure of the detection rule, it is preferable to work with a nonparametric regression architecture for $q(t, \mathbf{x})$. (For example a linear regression model for q would imply that $\mathfrak{S}$ in (5.4) is defined through linear constraints, i.e. it forms a simplex in $\mathcal{X}$.) In addition, nonparametric regression is typically more robust for dealing with the non-Gaussian residuals ϵ that arise in our model.

There are numerous nonparametric regression frameworks that can be used, including splines, Gaussian processes, or generalized additive models; see e.g. the classic monograph by Hastie et al. [2009]. Note that even though $\mathbf{x} \mapsto V(\mathbf{x})$ is continuous, some discontinuous response surfaces might also be helpful, such as random forests or dynamic trees [Gramacy and Ludkovski, 2015]. In the present thesis we take up a simple variant of splines, known as piecewise linear regression, as well as a variant of local linear regression, known as Loess.

Piecewise linear regression[Hastie et al., 2009, James et al., 2013] divides the data into regions and fits a polynomial regression in each region separately. While piecewise linear regression is not a nonparametric regression, it provides an easy and convenient basis for prediction and interpretation.

The piecewise linear response model with regions $\mathfrak{R}_j = \{\mathbf{x} : \mathbf{x} \in \mathfrak{R}_j\}, j = 1, \ldots, M$,

is of the form:

$$\hat{q}^{PL}(t, \mathbf{x}) = \sum_{i=1}^{r} \sum_{j=1}^{M} \hat{\beta}_{ij}(\mathbf{x}) B_i(\mathbf{x}) \mathbb{1}_{\{\mathfrak{R}_j\}}(\mathbf{x}), \tag{5.5}$$

where $B_i(\cdot)$ is the set of r pre-specified basis functions and $\hat{\beta}_{ij}$ are estimated least squares regression coefficients at $\mathbf{x}$ for region $\mathfrak{R}_j$, and $\mathbb{1}_{\{\mathfrak{R}_j\}}(\mathbf{x})$ is defined as

$$\mathbb{1}_{\{\mathfrak{R}_j\}}(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in \mathfrak{R}_j, \\ 0, & \text{if } \mathbf{x} \notin \mathfrak{R}_j. \end{cases} \tag{5.6}$$

Loess fits weighted linear regression models to localized subsets of data, determined using a kernel function, specifically a k-nearest-neighbor algorithm [Cleveland and Devlin, 1988]. Compared to classical linear models, Loess better handles outliers and heteroscedasticity, and also does not make assumptions about the global shape of the response surface.

The Loess response model is of the form

$$\hat{q}^{Loess}(t, \mathbf{x}) = \sum_{i=1}^{r} \hat{\beta}_i(\mathbf{x}) B_i(\mathbf{x}), \tag{5.7}$$

where $B_i(\cdot)$ is the set of r pre-specified basis functions and $\hat{\beta}_i$ are the estimated regression coefficients at $\mathbf{x}$. Given input matrix $\vec{X}$ and matching response vector Q, $\hat{\beta}$ is fitted using local least-squares minimization

$$\hat{\beta}(\mathbf{x}) := \arg\min_{\vec{\beta} \in \mathbb{R}^r} K_\lambda(\mathbf{x}, \vec{X})(Q - B(\vec{X})^T \vec{\beta})^2, \tag{5.8}$$

where $K_\lambda(\mathbf{x}, \vec{X})$ is the weighting kernel. The idea behind the kernel is to base the predicted $\hat{q}(t, \mathbf{x})$ on the samples in the neighborhood of $\mathbf{x}$, weighted by their distance

from $\mathbf{x}$ [Hastie et al., 2009, Sec. 2.8.2]. The size of the neighborhood is controlled by the smoothing parameter λ. If $\lambda < 1$, only a proportion λ of the samples will be used in fitting. The smaller λ, the more "wiggly" the fit $\hat{q}(t, \cdot)$ is going to be since fewer samples are used in computing $\hat{\beta}(\mathbf{x})$. Loess can be viewed as a special kernel regression method, with the prediction being a weighted average of the responses q^n: $\hat{q}(t, \mathbf{x}) = \sum_n l_n(\mathbf{x}) q^n$ for the equivalent kernel $l(\cdot)$. In our numerical examples, we use the implementation of Loess provided in the `R` by the built-in package `stats` R Core Team [2017], which uses a tri-cubic kernel and linear, first-order basis functions; the smoothing parameter used is $\lambda = 0.4$.

5.3 Experimental Design

The aim of the response surface is to maximize the accuracy of $\hat{\mathfrak{S}}_t$. This is equivalent to maximizing model fidelity along the boundary of the detection map. Statistically, for a localized response surface, accuracy is primarily driven by the local density of the input data that is specified by the experimental design $\mathcal{Z}$. Hence, to maximize our confidence regarding the boundary of $\mathfrak{S}_t$ in (5.4), we generate appropriate, adaptively chosen experimental designs $\mathcal{Z}$. This is achieved using the Sequential RMC (SRMC) framework introduced by Gramacy and Ludkovski [2015]. SRMC uses tools from active learning/Bayesian optimization to gradually *grow* the design $\mathcal{Z}$ so as to zoom-in to the boundary of $\hat{\mathfrak{S}}_t$. This is done by first quantifying the accuracy of the existing response surface and then adding new design sites so as to maximize information gain. See Gramacy and Ludkovski [2015], Hu and Ludkovski [2017] for details. The SRMC approach is illustrated in Figure 6.1 where the adaptively generated experimental design $\mathcal{Z}$ (of size 2000 in the figure) is highly concentrated around the detection boundary $\partial\mathfrak{S}$. This targeted sampling of outbreak scenarios allows for more efficient estimation, in particular

lowering the local standard errors $\hat{v}(\mathbf{x})$ along $\partial\hat{\mathfrak{S}}_t$, cf. the right panel of Figure 6.1.

In (5.4) the boundary of $\hat{\mathfrak{S}}_t$ corresponds to the regions of $\mathcal{X}$ where the cost difference between immediate detection and waiting is zero. Hence, we aim to have more design points in regions where $\{\hat{q}(t,\mathbf{x}) \simeq 0\}$. To this end, we define the "posterior" measure of response surface accuracy via

$$p(\mathbf{x}) := \Phi\left(\frac{-|\hat{q}(t,\mathbf{x})|}{\sqrt{\hat{v}(\mathbf{x})}}\right), \tag{5.9}$$

where Φ is the standard normal cdf and the predictive variance $\hat{v}$ measures the standard error of the surrogate prediction

$$\hat{v}(\mathbf{x}) = \hat{\sigma}^2(\mathbf{x})\|l(\mathbf{x})\|^2, \tag{5.10}$$

with $\hat{\sigma}^2(\mathbf{x})$ the estimated variance of ϵ around $\mathbf{x}$ in (5.3) and $l(\cdot)$ is the kernel defined in Section 5.2, see [Hastie et al., 2009, Sec 6.1.2].

The motivation for (5.9) is that $p(\mathbf{x})$ mimics the Bayesian posterior probability of estimating the wrong *sign* (conditional on the samples in $\mathcal{Z}$) of $q(t,\mathbf{x})$, assuming that the posterior distribution is Gaussian with the empirical mean $\hat{q}(t,\mathbf{x})$ and variance $\hat{v}(\mathbf{x})$.

The defined metric $p(\cdot)$ serves as a guide to augment new design locations. Namely, it defines an acquisition function $w(\mathbf{x})$ for greedily growing $\mathcal{Z}$, similar to active learning methods MacKay [1992]. The acquisition function is highest in the regions where $p(\mathbf{x})$ is close to 0.5, which correspond to $\partial\hat{\mathfrak{S}}_t$. Our main choice is

$$w^{\min}(\mathbf{x}) = \min\left[p(\mathbf{x}), 1-p(\mathbf{x})\right]. \tag{5.11}$$

Alternatives include the Gini weights $w^{\text{gini}}(\mathbf{x}) = p(\mathbf{x})\left(1-p(\mathbf{x})\right)$ and Entropic weights $w^{\text{Ent}}(\mathbf{x}) = -p(\mathbf{x})\log p(\mathbf{x}) - (1-p(\mathbf{x}))\log(1-p(\mathbf{x}))$.

To speed up the response surface modeling, which requires refitting of $\hat{q}(t,\cdot)$ multiple times, we used batch steps, incrementally working with designs $\mathcal{Z}^{(N)}$ of size $N = N_0, N_0+N', \ldots, N^{end}$. At each sequential design iteration, an additional N' design points $\{\mathbf{x}_0^n\}_{n=N+1}^{N+N'}$ are added to existing $\mathcal{Z}^{(N)}$. Those are sampled multinomially in proportion to the acquisition function $w(\cdot)$ from a candidate set X_{finite}. Both the initial design $\mathcal{Z}^{(N_0)}$ and the candidate sets X_{finite} are generated using Latin hypercube sampling (LHS) of size D from $\mathcal{X}$. The overall procedure, summarized in Algorithm B.6, finally refits at each iteration the Loess model for $\hat{q}$ (and hence $\mathfrak{S}_t$), grows the experimental design $\mathcal{Z}^{(N+N')} = \mathcal{Z}^{(N)} \cup \{\mathbf{x}_0^n\}_{n=N+1}^{N+N'}$, and recomputes the acquisition function (5.11). As the design size gets larger, we expect that the implied empirical estimate $\partial\hat{\mathfrak{S}}_t^{(N)}$ gets closer to the true $\partial\mathfrak{S}_t$.

Remark 5 *One can apply standard, non-sequential RMC by skipping the inner while loop (steps 7-15) in Algorithm B.6. This reduces to building a response model on a pre-specified (possibly randomized) design* $\mathcal{Z} := \{\mathbf{x}_0^n\}_{n=1}^{N_0}$*, keeping all other steps as is.*

Chapter 6

Case Studies

To illustrate the dynamic detection strategy within our 2-pool model, we present two case studies in this Chapter. In Section 6.1 we will start with the simple case, where we assume that some disease follows a two-population SIR model (discussed in Chapter 3) and therefore we could use the First Reduced Model (discussed in Section 4.1.1) for this detection strategy. In Section 6.2 we will recreate the measles epidemic in the UK using the data we discussed in Section 3.3 and use the Second Reduced Model (discussed in Section 4.2.1) for the detection strategy.

6.1 Case Study Using the First Reduced Model

Table 6.1 summarizes the parameters used for this case study. Epidemic parameters are taken to be $\mu_{SI}(t) = 0.75$ and $\mu_{IR}(t) = 0.5\ \forall t$. Thus, the initial reproduction ratio is $\mathcal{R}_0 = \mu_{SI}(t)/\mu_{IR}(t) = 1.5$, which is a moderately infectious epidemic. We assume that the pool mixing parameter is $\gamma = 0.01$, which is reasonable for pools representing well-separated cities or counties. The inference noise in (4.3) is taken to be Gaussian with variance $\delta_t \sim \mathcal{N}(0, \sigma_\delta^2 = 0.01^2)$. For the detection costs in (4.4)-(4.5), we take,

without loss of generality $C_{\text{Delay}} = 1$ and fix $C_{\text{FA}} = 20$. As we will see, this corresponds to a moderate penalty for false alarms.

For the detection map in Figure 6.1 in Section 6.1 we used an initial design of $N_0 = 200$, which was grown over 10 iterations with $N' = 200$ to a final design of $N^{end} = 2000$. The acquisition function was $w^{\min}$ and the candidate sets X_{finite} of size $D = 2500$ were generated with LHS. Since detection happens while $I^{(1)}$ is still relatively small, we restricted the response surface regression domain to $I^{(1)} \in \{0, 1, \ldots, 400\}, S^{(1)} \in \{1000, \ldots, 2000\}$. Lastly we note that the method is still computationally intensive, with the bulk of the effort spent on generating $T \cdot N^{end}$ scenarios of $\mathfrak{X}$; running times (on a 8-core 2.27GHz machine with 12GB of RAM) were about 20 minutes.

So far the case study features a three-dimensional state $\{S^{(1)}, I^{(1)}, P\}$, so that the resulting detection maps are in 3-D. To aid visualization, we consider a variant with a reduced dimension. Namely, we drop the component $S^{(1)}$ measuring the number of infecteds in Pool 1. Indeed, at the early stages of the outbreak the ratio $S_t^{(1)}/M^{(1)}$ is approximately one. As a result, one may assume that the rate of infections in Pool 1 is simply $\beta I_t^{(1)}$, which corresponds to the classical branching process epidemic model [Andersson and Britton, 2000, Ch. 6]. It is known [Ball and Donnelly, 1995] that this approximation remains valid up to $t = O(\log(M^{(1)})$ by which time, $I_t^{(1)} = O(\sqrt{M^{(1)}})$; therefore it works especially well in large populations, and hence is termed the large-population (LP) approximation. The LP model only has two dimensions, $\mathfrak{X}' := \{I^{(1)}, P\}$ allowing us to plot the corresponding 2-D stopping set $\mathfrak{S}^{LP}$.

Epidemic:	$M^{(1)} = 2000$	$S_0^{(1)} = M^{(1)} - I_0^{(1)}$	$\sigma_\delta = 1/100$
	$\mu_{SI}(t) = 0.75$	$\gamma = 0.01$	$\mu_{IR}(t) = 0.5$
Costs/Penalties:	$C_{\text{FA}} = 20$	$C_{\text{Delay}} = 1$	

Table 6.1: Outbreak and costs parameters for the case study described in Section 6.1. The parameter σ_δ refers to the noise in P, cf. (4.3).

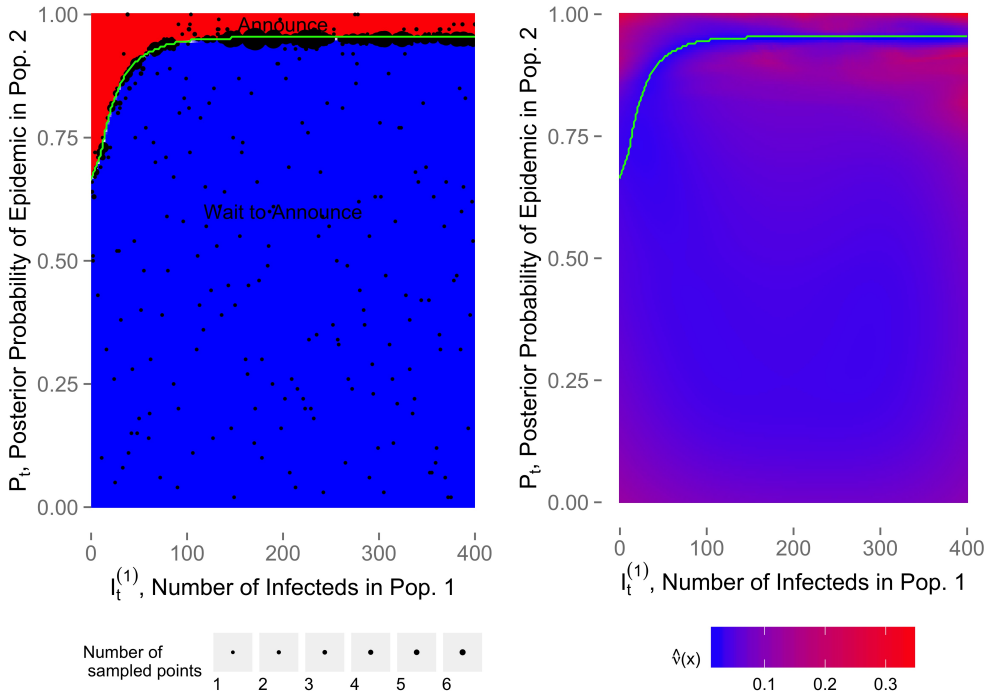


Figure 6.1: The left panel: detection rule $\mathfrak{S}_{20}^{LP}$ in terms of $I^{(1)}$ and P. The detection boundary $\partial\mathfrak{S}_{20}^{LP}$ is shown with the solid curve. We also show the experimental design $\mathcal{Z}$ that was used, illustrated with the scatterplot. The size of the pixel corresponds to the number of times that neighborhood was sampled. The right panel: standard errors $\hat{v}(\mathbf{x})$ from (5.10). Observe lower standard errors in regions where the design $\mathcal{Z}$ is more dense.

Figure 6.1 shows $\mathfrak{S}^{LP}$ generated under the conditions of Table 6.1 along with the above large population assumption. As expected, epidemic detection is triggered once the posterior probability P_t of $\{I_t^{(2)} > 0\}$ is high enough. However, we observe that detection is also highly sensitive to values of $I_t^{(1)}$; for instance detection is progressively delayed as $I_t^{(1)}$ gets bigger. This dependence between the two pools in terms of decision making illustrates the underlying cross-pool information fusion. Intuitively, detection should take place once P_t is high enough. However, conditional on a fixed P_t, the larger number of Pool 1 infecteds makes an impending outbreak in Pool 2 more likely, lowering waiting costs. Mathematically, recall that in (4.3), the growth rate of P increases in $I^{(1)}$.

As a result, for large values of $I_t^{(1)}$, one may expect that the next-stage P_{t+1} will also be large, i.e. move into the "Announce" region quicker. This again lowers the waiting costs and, therefore, delays announcement.

6.1.1 Evaluating Detection Rules

Figure 6.2 shows dynamic decision-making in the LP model through a collection of generated trajectories of $\mathfrak{X}' = \{I_t^{(1)}, P_t\}$ and their corresponding detection times τ^{LP}, the first time the state process $\mathfrak{X}'$ enters the stopping set $\mathfrak{S}^{LP}$. We observe that the trajectories generally move north-east, as both P and $I^{(1)}$ tend to increase. However, the rate at which they grow and the precise direction are uncertain and vary across scenarios. Consequently, at detection, both $P_{\tau^{LP}}$ and $I_{\tau^{LP}}^{(1)}$ have a nontrivial distribution.

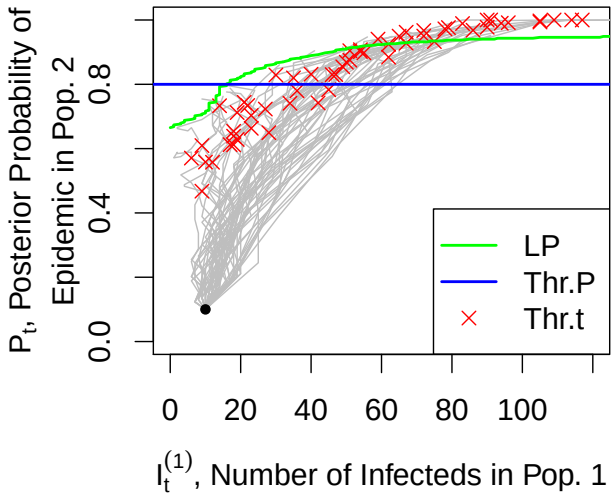


Figure 6.2: Fifty sampled epidemic trajectories $\{I_t^{(1)}, P_t\}, t = 1, \ldots, \tau$ emanating from the initial state $I_0^{(1)} = 10$ and $P_0 = 0.1$. We show the LP detection boundary (namely $\partial\mathfrak{S}_{20}^{LP}$), as well as a threshold strategy that announces the epidemic as soon as $P_t \geq \bar{P} = 0.8$. Lastly, the red crosses denote the locations of the trajectories at $t = 8$, which is the basis of the alternate Threshold-t strategy.

To better understand the detection map $\mathfrak{S}^{LP}$, we analyze the resulting detection

strategy given by τ^{LP} and compare it to alternatives. Two classes of simpler detection rules are Threshold-P and Threshold-t. The Threshold-P strategy announces an outbreak as soon as $P_t \geq \bar{P}$ for a given threshold $\bar{P}$. Hence, it acts solely based on local (posterior) information about Pool 2. This mimics the CDC policy [Hutwagner et al., 2005] of announcing an epidemic when the number of infecteds in Population 2 crosses some pre-specified level. In contrast to the fused detection strategy with a curved detection boundary which jointly takes into account both P_t and $I_t^{(1)}$, Threshold-P rule only uses P_t for detection decisions, yielding a flat, horizontal detection boundary in Figure 6.2. The threshold-t strategy is a simple non-adaptive strategy that announces at the fixed stage $\bar{t}$. It is illustrated in Figure 6.2 where we record the joint distribution of $I_{\bar{t}}^{(1)}, P_{\bar{t}}$ at $\bar{t} = 8$.

	Detection time τ		Cost $c(\mathfrak{X}_{0:\tau^{(t)}})$		PFA $E[1 - P_\tau]$
	Mean	StDev.	Mean	StDev.	
Optimal	8.86	2.59	6.53	1.70	8.2%
LP	9.32	2.95	6.57	1.81	6.4%
Threshold-P	7.88	2.85	7.03	1.58	15.3%
Threshold-t	8.00	N/A	7.18	2.21	14.4%

Table 6.2: Comparison of Optimal, Large Population(LP), Threshold-P with $\bar{P} = 0.8$ and Threshold-t with $\bar{t} = 8$ strategies. Statistics are based on 1000 synthetic trajectories of $\{S^{(1)}, I^{(1)}, P\}$ with the starting value {1990, 10, 0.1}.

Returning to the full 3-D model with state $\mathfrak{X}$, we evaluate the resulting optimal detection strategy τ^* and proceed to compare its performance against the other potential detection rules discussed above. Specifically, the first two alternatives are a Threshold-P rule with $\bar{P} = 0.8$ (declare an epidemic if its probability is above 80%) and a Threshold-t strategy with $\bar{t} = 8$. The latter was found to be the best strategy among those that declare outbreak at a fixed stage. The last alternative is the LP strategy τ^{LP} from last section. Recall that τ^{LP} makes decisions while ignoring $S^{(1)}$. In that sense, when applied to the full 3-D model, it gives a simplified, but still adaptive, detection rule. To recap,

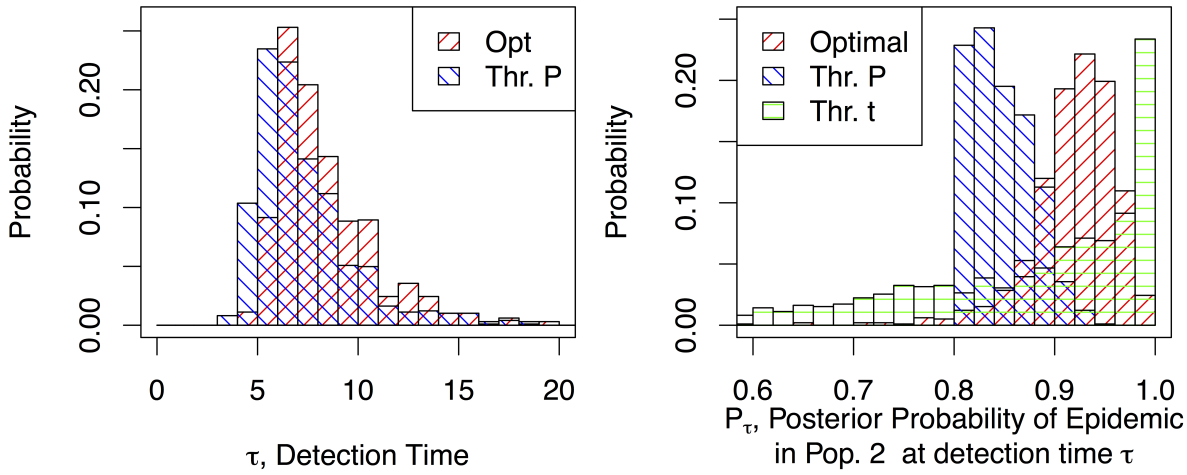


Figure 6.3: Summary statistics of different detection strategies constructed from 1000 sample epidemic trajectories. The LP detection strategy is from Figure 6.1. Right: Distribution of detection times τ; Left: Distribution of posterior probability of outbreak in Pool 2 at detection time, P_τ.

the Threshold-t strategy is completely non-adaptive; Threshold-P only relies on P_t; LP relies on $\{I_t^{(1)}, P_t\}$; the Optimal strategy uses all of $\{S_t^{(1)}, I_t^{(1)}, P_t\}$.

To compare the performance of the above competing strategies, we fix the initial condition at $S_0^{(1)} = 1990, I_0^{(1)} = 10$, and $P_0 = 0.1$, so that there are 10 infecteds in Pool 1 and 10% prior probability of the epidemic already in Pool 2. Then we simulate 1000 epidemic trajectories $\{\mathbf{x}_{0:\tau}^n\}$, $n = 1, \ldots, 1000$, emanating from this fixed initial condition up to the detection time τ (which depends in turn on the strategy used). Table 6.2 then presents the resulting summary statistics based on these frozen 1000 trajectories (note that there are no analytic formulas to obtain these metrics, so we must resort to simulation).

The comparison is done in terms of several different metrics, including detection costs $c(\mathfrak{X}_{0:\tau^{(t)}})$, distribution of detection times τ, and frequency of false alarms, represented by $d(\mathfrak{X}_\tau) = 1 - P_\tau$ in our setup. As expected, the Optimal strategy with detection

time τ^* that directly optimizes the cost-benefit in the full model performs best. The corresponding expected costs are $V(\mathbf{x}_0) \simeq 6.53$, with average detection time $E[\tau^*] \simeq 8.86$. It outperforms the Threshold-P strategy by about 7% in terms of reducing detection costs, and the Threshold-t strategy by about 9%. These are nontrivial cost savings which highlight the benefit of information fusion. Table 6.2 also shows that the 2-D LP approximation performs well in this example, generating very similar expected costs. At least for this case study, detection happens early enough that the branching process approximation of the outbreak works fine.

Recall that our model is stochastic and generates adaptive detection strategy. Hence the detection time τ^* is a random variable. As shown in Table 6.2, the corresponding standard deviation $StDev(\tau^*) \simeq 2.6$ is substantial. This illustrates the sub-optimality of the Threshold-t strategy that stops at a fixed $\bar{t}$ with $StDev(\bar{t}) = 0$ trivially. Not surprisingly, the ability to delay or speed up outbreak announcements based on latest data are crucial for optimizing policy making. We also note that compared to the Threshold-P strategy, the Optimal strategy tends to announce later, $E[\tau^*] \simeq 8.86 > 7.88 \simeq E[\tau^{Thr-P}]$. This is also confirmed by the respective histograms of τ^* and τ^{Thr-P} as shown in Figure 6.3. However, we emphasize that the detection rules do not have a clear ordering. In other words, the random variables τ^*, τ^{Thr-P}, etc., cannot be directly compared.

A complementary metric of detection quality is provided by the probability of false alarms, $PFA := E[1 - P_\tau]$. For the optimal strategy we find that $PFA^* = 8.2\%$. In contrast, for Threshold-P strategy, we have $PFA^{Thr-P} = 15.3\%$. Note that because we use a discrete-time model, time of detection P_τ will strictly exceed the threshold $\bar{P} = 0.8$, hence $PFA^{Thr-P} < 1 - \bar{P}$. The histograms of P_τ are shown in Figure 6.3 and confirm the qualitative difference among the detection strategies. The Threshold-P strategy only stops once $P_t > \bar{P}$ so that P_τ has support on roughly $[0.8, 0.9]$. In contrast, the adaptive Optimal (and LP) strategies, have a much wider range for P_τ. In particular, sometimes

epidemics are announced even before P_t hits the level 0.8.

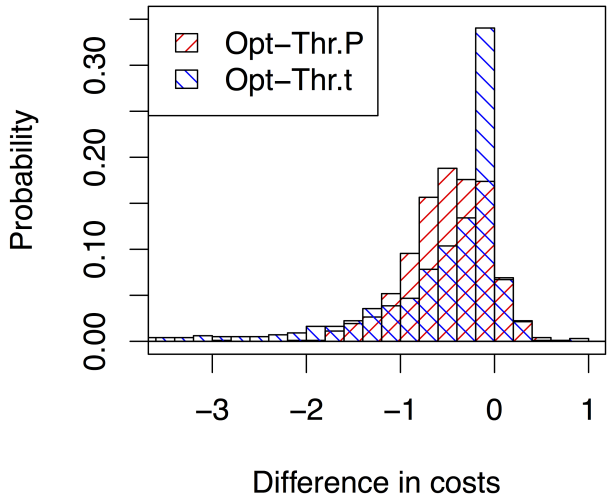


Figure 6.4: Relative detection costs across different strategies. The histogram shows the distribution of the difference in costs along the 1000 simulated trajectories, namely $c(\mathbf{x}_{0:\tau^*}) - c(\mathbf{x}_{0:\tau^{Thr-P}})$, and $c(\mathbf{x}_{0:\tau^*}) - c(\mathbf{x}_{0:\tau^{Thr-t}})$.

To further quantify the improvement provided by the Optimal detection rule, Figure 6.4 gives a scenario-by-scenario comparison of realized detection costs. Note that in hindsight, τ^* may sometimes perform worse that τ^{Thr-P} or even τ^{Thr-t}. Figure 6.3 plots the histogram of the difference in costs for each trajectory $\mathbf{x}^n_{0:t}$, $n = 1, \ldots, 1000$, namely $c(\mathbf{x}_{0:\tau^*})$, $c(\mathbf{x}_{0:\tau^{Thr-P}})$, and $c(\mathbf{x}_{0:\tau^{Thr-t}})$. We find that the costs computed with Optimal/LP strategies are smaller than costs computed with Threshold strategies for more than 80% of the trajectories.

To sum up, we observe material improvement when using the Optimal detection rule in this case study. Moreover, the obtained detection rule is substantially different from the thresholding protocol. On the one hand, the adaptive detection time τ^* exhibits a wide spread and is highly non-constant across trajectories. On the other hand, the posterior probability of false alarms P_{τ^*} is also strongly variable. As a result, the average

frequency of false alarms is drastically lowered relative to Threshold-P strategy, reducing overall expected costs.

6.1.2 Effect of Detection Cost Parameters

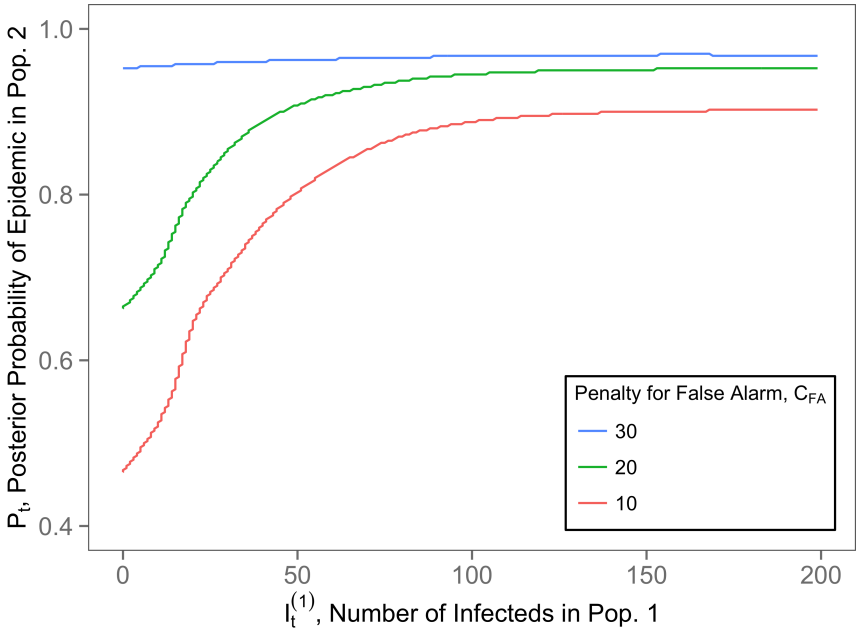


Figure 6.5: Boundaries of detection maps $\partial\mathfrak{S}_{20}^{LP}$ constructed based on different penalties for false alarm, C_{FA}.

C_{FA}	τ^*		Cost		$PFA = E[1 - P_{\tau^*}]$
	Mean	StDev.	Mean	StDev.	
10	6.84	1.62	5.32	0.99	21.4%
20	8.87	2.60	6.54	1.71	8.3%
30	9.61	2.79	7.21	2.22	5.3%

Table 6.3: Summary statistics of the Optimal detection strategy τ^* for different false alarm penalties C_{FA}. Statistics are based on 1000 synthetic trajectories of $\{S^{(1)}, I^{(1)}, P\}$ with the starting value {1990, 10, 0.01}.

The main parameter in our quickest detection setup is the ratio of the cost of false alarms and the cost of detection delay, $C_{\text{FA}}/C_{\text{Delay}}$. A high ratio penalizes premature

announcements and requires more care in the assessment of the potential outbreak in Pool 2. A low ratio invites more aggressive actions. To better understand the role of this ratio, we refer to Figure 6.5, where we show several detection boundaries $\partial\mathfrak{S}^{LP}$ corresponding to varying C_{FA}, while $C_{\text{Delay}} = 1$ is kept fixed. As expected, a lower C_{FA} enlarges the Announce set $\mathfrak{S}$. In particular, the boundary $\partial\mathfrak{S}$ shifts down and to the right. As a result, starting from a fixed location $(I_0^{(1)}, P_0)$, the stopping set $\mathfrak{S}$ will be reached sooner, so that τ decreases (in the sense of stochastic dominance for the corresponding random variables). This is confirmed in Table 6.3 that reports statistics for τ^* and various C_{FA}. We find that $E[\tau^*] = 8.86$ when $C_{\text{FA}} = 20$, but is only $E[\tau^*] = 6.84$ for $C_{\text{FA}} = 10$. Simultaneously, the frequency of premature announcements PFA will increase. The precise relationship is however nonlinear. Lowering C_{FA} from 20 to 10, the PFA rises dramatically to about 21% from 8%. Conversely, raising C_{FA} to 30 only reduces PFA to 5.3%.

6.2 UK Case Study Using the Second Reduced Model

We set up a detection strategy for announcing a measles epidemic in Bristol using the information about the epidemic in London. The parameters of epidemic for these two cities are presented in Table 3.1. We assume that mortality rate $\mu_B^{(1)}(t) = \mu_B^{(2)}(t) = \mu_D^{(1)}(t) = \mu_D^{(1)}(t) = 3.836 \cdot 10^{-4}$ per week, the mean contact rate $\bar{\mu}_{SI}^{(1)} = \bar{\mu}_{SI}^{(2)} = 4.75$ per week, and the pool mixing parameter $\gamma = 0.02$. We use Two Population Asymmetrical epidemic model (see Section 3.2) for modeling this epidemic.

We start by creating a database of 10000 possible epidemic trajectories simulated with Two Population Asymmetrical epidemic model as well as the trajectories of $\{\mu_t, \sigma_t^2\}$ from a 'Poisson' algorithm for 20 years. We remove the burn-in period of 5 years. There are two reasons for creating this database. First, we can compute $\eta^{(2)}$, the average proportion

of susceptible individuals in the second population for each week of the year $w = t \mod 52$, which is used in Algorithm B.4. Second, the database has 150000 locations of $\left(w, S_0^{(1)}, E_0^{(1)}, I_0^{(1)}, R_0^{(1)}, \mu_0, \sigma_0^2\right)$ (we drop the information about the second epidemic $X_t^{(2)}$) that will be used for sampling an initial design $\{\mathbf{x}_n\}_{n=1}^{N_0}$ for Algorithm B.6.

Based on the initial design we can propagate the trajectories and compute the cost difference via Algorithm B.5 (step 4 of Algorithm B.6). To generate N_0 trajectories of $\{S_t^{(1)}, E_t^{(1)}, I_t^{(1)}, R_t^{(1)}, S_t^{(2)}, I_t^{(2)}, R_t^{(2)}\}_{t=0}^{\min(\tau,52)}$ using the Two Population Asymmetrical SEIR-SIR epidemic model (see Section 3.2), we also need the starting points of the second population. We sample an approximation of $I_0^{(2)}$ from Gamma posterior distribution with parameters (α_0, β_0), which can be found via Equation (4.30) and (μ_0, σ_0^2). We approximate $S_0^{(2)}$ by multiplying the population size by $\eta^{(2)}$ and get $R_0^{(2)}$ by subtraction. Then we generate the trajectories of (μ_t, σ_t^2) using the Second Reduced Model (see Section 4.2.1). Therefore, for our design $\{\mathbf{x}_0^n\}_{n=1}^{N_0}$ we get the corresponding path-wise cost difference $\{q^n\}_{n=1}^{N_0}$, where we take $C_{\text{Delay}} = 1$ and fix $C_{\text{FA}} = 10$ for the detection costs in (4.34)-(4.35).

Therefore we have an 7-dimensional design $\{\mathbf{x}_0^n\}_{n=1}^{N_0}$ to regress against the path-wise cost differences $\{q^n\}_{n=1}^{N_0}$ in Equation (5.3). We simplify our design to 2 dimensions (w, μ) and use the piecewise linear model. We break weeks w into the two types of regions: $W_1 = \{w : w = 51, 52, 1, 14, 15, 16, 28, 29, 30, 31, 32, 33, 34, 35, 36, 42, 43, 44\}$ – the weeks that correspond to the school status 'On a Break' and $W_2 := W_1^c$ – the weeks that correspond to the school status 'In Session' . For the breakpoints of μ we compute the values of the three quartiles (25th, 50th, 75th percentiles respectively) of the current design as well as $\bar{I}$, the inital threshold level for epidemic. Thus, for μ we get 4 ordered breakpoints m_i, $i = 1, 2, 3, 4$, giving us 5 regions for μ: $M_1 = \{\mu : \mu \leq m_1\}$, $M_2 = \{\mu : m_1 < \mu \leq m_2\}$, $M_3 = \{\mu : m_2 < \mu \leq m_3\}$, $M_4 = \{\mu : m_3 < \mu \leq m_4\}$, $M_5 = \{\mu : \mu > m_4\}$. Therefore we have a total of 10 regions: $\mathfrak{R}_{kl} = \{(\mu, w) : \mu \in M_k, w \in W_l\}$, $k = 1, 2, \ldots, 5$, $l = 1, 2$.

Piecewise linear model (in comparison to loess) allows us to add trigonometric terms $\sin(2\pi w/52) + \cos(2\pi w/52)$ to handle the weeks of the year term, therefore making sure that the regression fit at week 0 matches the one of week 52 [Montgomery et al., 2015]. We also add a quadratic term in μ as well as the interaction terms, and therefore our final regression model has 23 coefficients:

$$\hat{q}^{PL} = \hat{\beta}_0 + \hat{\beta}_1 \sin\left(\frac{2\pi w}{52}\right) + \hat{\beta}_2 \cos\left(\frac{2\pi w}{52}\right) + \sum_{i=1}^{2}\sum_{k=1}^{5}\sum_{l=1}^{2} \hat{\beta}_{ij} B_i(\mu) \mathbb{1}_{\{\mathfrak{R}_{kl}\}}((\mu, w)), \tag{6.1}$$

where the polynomial basis functions are $B_i(\mu) = \mu^i$, $i = 1, 2$ and the indicator function $\mathbb{1}_{\{\mathfrak{R}_{kl}\}}((\mu, w))$ was defined in Equation (5.6).

Unfortunately the decrease in dimensions means we can not straightforwardly simulate the behavior of $\left(S_t^{(1)}, E_t^{(1)}, I_t^{(1)}, R_t^{(1)}\right)$. SRMC approach adds additional observations to the boundary $\partial\hat{\mathfrak{S}}_t$. To get these observations, we need to simulate the trajectory from $\left(w, S^{(1)}, E^{(1)}, I^{(1)}, R^{(1)}, \mu, \sigma^2\right)$. The coordinates of our boundary are just (w, μ_t) with unknown position of epidemic in the first Pool. Thus, we will use Regression Monte Carlo approach to get the detection map (see Algorithm B.6 with $N_0 = N^{end}$) and we leave an implementation of SRMC in this problem as future work.

Figure 6.6 presents the series of detection rules that we got using Regression Monte Carlo (see Section 5.1) without implementing Model Predictive Control (MPC). As it was shown in Section 5.1, to get the detection rule $\hat{\mathfrak{S}}_t$, we use detection rules $\hat{\mathfrak{S}}_{t-1:0}$ at times $0, 1, \ldots, t-1$ and $\mathfrak{S}_0$ at iterations $t+1, t+2, \ldots$. So we can observe rather flat boundary $\partial\hat{\mathfrak{S}}_1$ (first panel of Figure 6.6) because we only used our initial detection rule $\hat{\mathfrak{S}}_0$ at time $t = 0$. Then with each iteration our boundary changes: at iteration 14

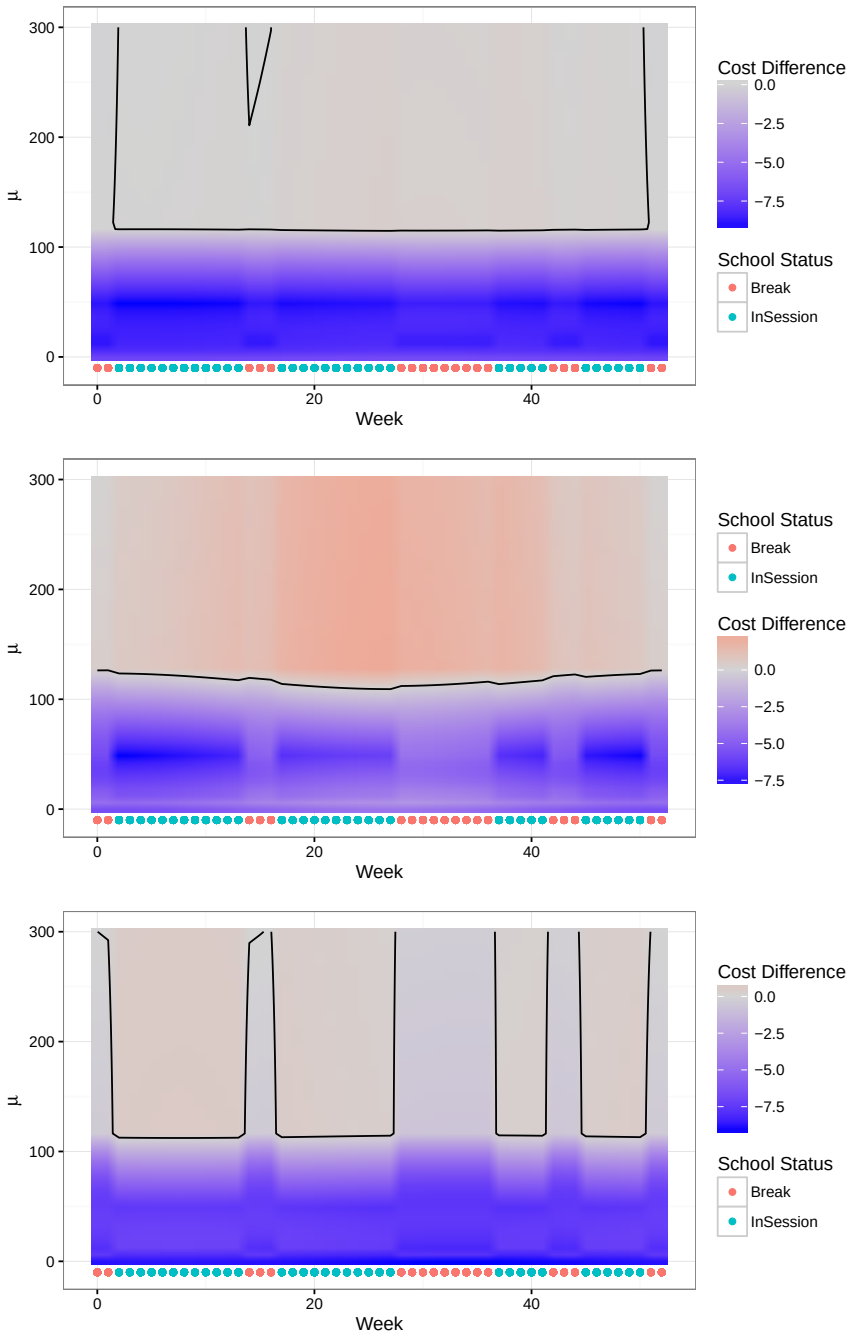


Figure 6.6: Detection Rules $\hat{\mathfrak{S}}_1$(first panel), $\hat{\mathfrak{S}}_{14}$ (second panel), $\hat{\mathfrak{S}}_{32}$ (third panel). The Detection Boundaries $\partial\hat{\mathfrak{S}}_1, \partial\hat{\mathfrak{S}}_{14}, \partial\hat{\mathfrak{S}}_{32}$ are shown with the solid curves on the respective plots. The regions with a red shade color correspond to the state space, where we announce the epidemic, while the regions with a blue color shade correspond to the state space where we do not announce an epidemic. Initial detection map we set $\mathfrak{S}_0 = \{\mathfrak{X} : \mu_t > 100\}$. Cost difference $\hat{q}$ is defined as difference between future and immediate costs.

we still observe rather flat boundary $\hat{\mathfrak{S}}_{14}$, however the small bumps around the school breaks appear, which means the epidemic's announcement happens for a bigger value of μ compared to the time when school is in session. Detection boundary $\hat{\mathfrak{S}}_{32}$ is even more prominent than the boundary $\hat{\mathfrak{S}}_{14}$: we do not announce epidemic during breaks.

6.2.1 Evaluating Detection Rules

For this section we assume that there is an epidemic if the number of infecteds crosses $\bar{I} = 50$. Our initial detection rule is a Threshold detection rule - to announce an epidemic if $\mu > \bar{I}$. We generate maps $\hat{\mathfrak{S}}_{1:10}$, and then employ Model Predictive Control (MPC) for 50 more iterations (MPC was discussed at the end of Section 5.1) to get $\hat{\mathfrak{S}}^{(11:60)}$. To evaluate those, we generate 1000 epidemic trajectories in two populations starting from Week 34, where the starting conditions were sampled from our database of 150000 locations with $w = 34$ defined in the beginning of Section 6.2, and lasting for one year. Based on those trajectories we can compute the Costs using Equation (2.3) and Probability of False Alarm (PFA) for each $\hat{\mathfrak{S}}_t,\ t = 1, \ldots, 60$, where

$$\text{PFA} = E(\Psi_{I^2_\tau}(\bar{I})). \tag{6.2}$$

Figure 6.7 shows how the costs and the probability of false alarm change with respect to the different time t (the Figure 6.7 also reflects how the Costs and the Probability of False Alarm are affected by the change of Cost of False Alarm C_{FA}, but we will discuss in Section 6.2.2). We see a steady decrease in both the Costs and the Probability of False Alarm, PFA. Therefore, the detection rule is an optimal one .

For example, if we fix the Cost of False Alarm C_{FA} at 30 , we see that the Cost decreased from 6.62 to 2.51 units, which is about a 62% decrease from the initial Threshold detection rule (if $\mu > 50$); and Probability of False Alarm decreased from 36.25% to

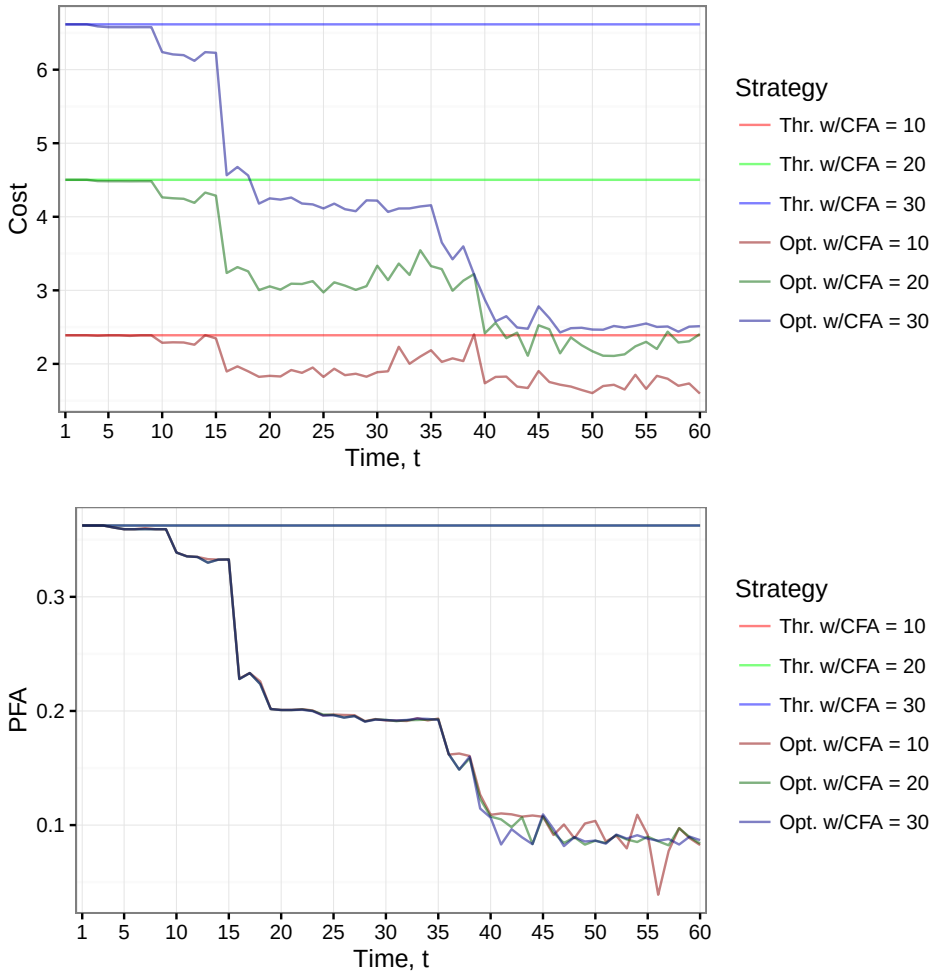


Figure 6.7: Relationship between estimated Optimal detection rules $\hat{\mathfrak{S}}_{1:60}$ and Threshold detection rules (initial detection map), shown via computation of Costs (top plot) and Probability of False Alarm (PFA) (bottom plot). The initial detection map - $\mathfrak{S}_0 = \{\mathfrak{X} : \mu_t > 50\}$

8.68%, which is a 76.1% decrease from the initial detection rule.

The mean of detection time τ of the 1000 trajectories increased slightly from 26.132 to 27.422 with standard deviations decreasing from 11.733 to 10.345. Detection time 26 correspond to week 11 of the year, which means that most of detection happens on average before the Spring Break. However the mean value of μ at detection time τ increased significantly from 68.94 (with standard deviation 19.3) to 103.56 (with standard deviation 15.5).

Thus, we observed the significant improvement in our Detection Rules compared to the initial one. These detection rules not only decreased the costs, but also reduced the probability of false alarm.

The final estimated detection map $\hat{\mathfrak{S}}_{60}$ for each of $C_{FA} = 10, 20, 30$ is given in Figure 6.8. As it was discussed before we do not announce an epidemic during the school breaks or do announce an epidemic for a higher level of μ during the school breaks.

6.2.2 Effect of Detection Cost Parameters

As it was seen in the previous case study in Section 6.1.2, the main parameter in our quickest detection setup is the ratio of the cost of false alarms and the cost of detection delay, $C_{\text{FA}}/C_{\text{Delay}}$. A high ratio penalizes premature announcements, while a low ratio invites more aggressive actions. As expected, a decrease in the cost of false alarm C_{FA}, when keeping the fixed cost of detection delay $C_{\text{Delay}} = 1$, caused the probability of false alarms (PFA) to drop. This is confirmed by observing the lines on Figure 6.7 or from Table 6.4 (while for the Threshold strategy this increase in the cost of false alarms didn't affect PFA). Notice that in Figure 6.8 the area, where we announce an epidemic, decreases as the Cost of False Alarm increases.

Table 6.4 also provides the Detection times τ, where detection time 26 corresponds

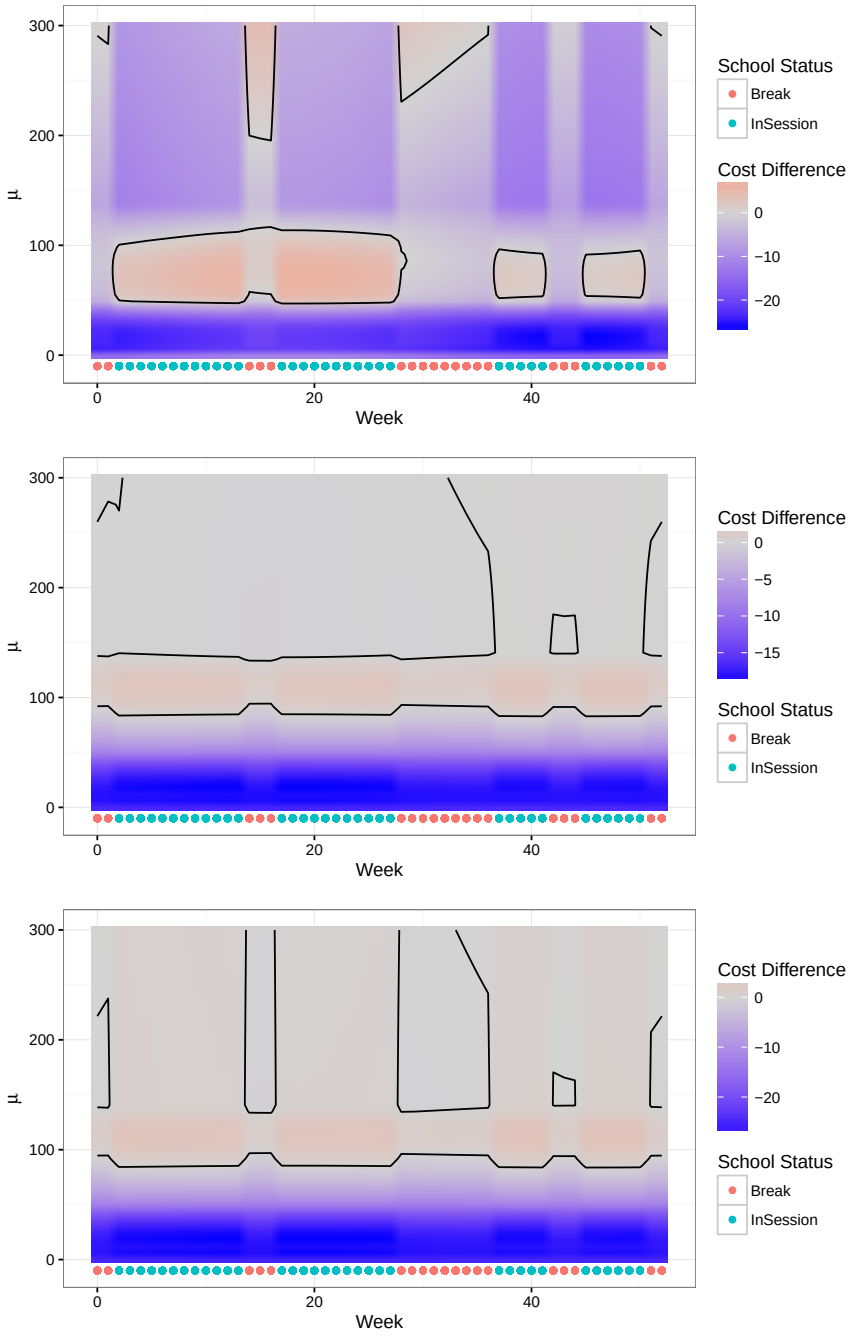


Figure 6.8: Detection Rules $\hat{\mathfrak{S}}_{60}$ computing using $C_{FA} = 10$ (top panel), $C_{FA} = 20$ (middle panel), $C_{FA} = 30$ (bottom panel). The Detection Boundaries $\partial\hat{\mathfrak{S}}_{60}$ are shown with the solid curves on the respective plots. The regions with a red shade color correspond to the state space, where we announce the epidemic, while the regions with a blue color shade correspond to the state space where we do not announce an epidemic.

to the 11th week of the year. The detection rule with $C_{\text{FA}} = 30$ announced an epidemic a little sooner than the detection rules with $C_{\text{FA}} = 10$ and $C_{\text{FA}} = 20$.

	C_{FA}	Detection time τ		Cost		PFA
		Mean	StDev.	Mean	StDev.	
Optimal	10	26.637	11.216	2.174	2.227	30.5%
Threshold	10	26.132	11.733	2.389	2.306	36.3%
Optimal	20	26.884	10.204	2.170	2.347	8.8%
Threshold	20	26.132	11.733	4.503	4.450	36.3%
Optimal	30	26.518	10.133	2.520	2.888	8.6%
Threshold	30	26.132	11.733	6.616	6.601	36.3%

Table 6.4: Comparison of Optimal strategy $\hat{\mathfrak{S}}_{60}$ with the initial detection rule $\mathfrak{S}_0 = \{\mathfrak{X} : \mu > 50\}$ and Threshold-I strategy with $\bar{I} = 50$. Statistics are based on 1000 synthetic trajectories of $\{\mu, w\}$.

Therefore, in both case studies we showed a decrease of Probability of False Alarm, which lowered the Costs, in comparison to Threshold strategies. We also varied a cost of false alarm C_{FA} in both case studies: the lower the C_{FA} enlarges the Announce set $\mathfrak{S}$ and therefore increases the Probability of False alarm. A common approach in the decision literature is to select a priori a desired level of PFA (say $PFA = 10\%$) and then numerically solve the inverse problem to obtain the corresponding C_{FA} and hence the corresponding detection rule $\mathfrak{S}$.

Chapter 7

Conclusion and Future Work

We have presented a framework for optimal detection of epidemics in a coupled meta-population model. Taking into account data about the second population is important because the epidemic in the original population can be detected more accurately. This is especially relevant when the original population is small and there is an epidemic in the nearby bigger population. In this scenario even a small number of Infected individuals in the original population is likely to produce an epidemic. Even if the populations are of the same size, if there is a big epidemic in the second population, the epidemic in the original population becomes more likely.

Our approach explicitly takes into account cost-benefit considerations regarding announcement of an epidemic, as well as spatial dependence across susceptible pools. Given the information about two populations and characteristics of the infection, our algorithm produces the full detection map which can then be used repeatedly for an announcement of epidemic. We demonstrate that information about the epidemic in one pool can be used to lower the detection costs in another pool, realizing savings compared to traditional threshold-type detection methods.

The presented SIR/SEIR framework gives a basic mechanistic description of disease

progression. More sophisticated versions might allow for age stratification, and heterogeneity among the meta-populations. One could also include further transitions beyond (3.1), such as immunity lapses $R^{(k)} \to S^{(k)}$ or vaccination $S^{(k)} \to R^{(k)}$.

Alternatively, one can also imagine more sophisticated models for the outbreak pseudo-posterior – recall that the proposed one was largely for convenience than any realism. For example, the Gaussian noise δ_t in the dynamics of P_t that was used in the case study may be better modeled via a Beta distribution (which arises naturally as conjugate to the Poisson increments of the fully-observed stochastic SIR model Lin and Ludkovski [2014]). Overall, the key requirement is the Markov structure which makes it possible to use regression against $\mathfrak{X}$ to describe the detection rule. The Markovianity requirement can be partly relaxed if one is willing to accept approximately-optimal solutions. Indeed, one can always project the optimal detection rule into the smaller space of rules that only depend on some subset $\mathfrak{X}'$; in other words restricting the detection map to only take into account some of the state-space dimensions. This idea was already discussed in Section 6.1 where we described the sub-optimal LP strategy.

Second, one may modify the cost structures (2.2)-(2.3) to better capture the desired detection goals. The presented costs were motivated by their classical analogues in sequential change-point detection, but might not be the most appropriate for public health contexts. The waiting cost in (2.3) was constant; it may be more realistic to make it proportional to $I_t^{(2)}$, which would correspond to fixed costs per infected.

For such more general formulations, the costs $d(\mathfrak{X})$ and $c(\mathfrak{X})$ would no longer be functions of pseudo-posterior distribution, and one would need to work with the full posterior distribution π_t of $I_t^{(2)}|\mathcal{G}_t$. The RMC framework could still be usable, namely we may use particle filtering (see Appendix A) [Ludkovski, 2009, Sheinson et al., 2014, Lin and Ludkovski, 2014], to obtain π_t along a simulated trajectory of the underlying epidemic model. Certainly, particle filtering can become computationally expensive,

making efficient inference essential. The integrated sequential inference plus optimization model would then allow to treat a partially observed version of a K-pool SIR model, where $K > 2$, and ultimately a larger-scale setup such as influenza surveillance across all 50 states, cf. Figure 1.1.

Appendix A

Particle Filtering

In this chapter we develop an alternative observation model based on observable infections in both populations and then present a Particle Filtering approach to solving our inference problem based on this model. Note that the notation of this section is slightly different from the rest of the thesis.

A.1 Observed Process and Noise

Suppose the observations are collected with fixed period τ. While collecting the observations (for example, observed cases $C^{(obs(2)}$ as in Section 4.2) we get into two problems. First, not all individuals might be diagnosed with the infection due to personal reasons. Second, false positive observations might be collected due to wrong diagnosis.

The solution to the first problem is the binomial sampling. Suppose we observe each infection with probability $p := 1 - \rho$ in a time period of length τ. Define $\mathbf{O}_{j\tau} = \{O_{j\tau}^{(1)}, \ldots, O_{j\tau}^{(K)}\}$, $j = 0, 1, 2, 3, \ldots$, to be a K-dimensional discrete time increasing jump process, which represents the cumulative number of new infections observed in each population by time $j\tau$. Thus, the increments of $\mathbf{O}_{j\tau}$ are defined as $\Delta_{j\tau}^{\mathbf{O}} = \mathbf{O}_{(j+1)\tau} - \mathbf{O}_{j\tau}$.

If we denote the new infecteds $\mathbf{S}_{(j+1)\tau} - \mathbf{S}_{j\tau}$ as $\Delta^I_{j\tau}$, then, by construction, the conditional density

$$\Delta^{\mathbf{O}}_{j\tau} | \Delta^I_{j\tau} \sim \text{Bin}(\Delta^I_{j\tau}, p) \tag{A.1}$$

for any $j = 0, 1, 2, 3, \ldots$.

The problem of false positives is solved by introducing an independent K-dimensional continuous time process $\mathbf{N}_{j\tau} = \{N^{(1)}_{j\tau}, \ldots, N^{(K)}_{j\tau}\}$, where

$$\Delta^{\mathbf{N}}_{j\tau} = \mathbf{N}_{(j+1)\tau} - \mathbf{N}_{j\tau} \sim \text{Poisson}(\lambda_0 \tau) \tag{A.2}$$

for any $j = 0, 1, 2, 3, \ldots$. Thus, $\mathbf{N}_{j\tau}$ represent the cumulative number of noisy observations in each population.

Define $\mathbf{Y}_{j\tau} = \mathbf{O}_{j\tau} + \mathbf{N}_{j\tau}$, which is a K-dimensional discrete time process for any $j = 0, 1, 2, 3, \ldots$. Thus, $\mathbf{Y}_{j\tau}$ correspond to the cumulative number of individuals diagnosed with this infection and we can define increments of $\mathbf{Y}_{j\tau}$ to be

$$\Delta^Y_{j\tau} = \mathbf{Y}_{(j+1)\tau} - \mathbf{Y}_{j\tau} = \Delta^{\mathbf{O}}_{j\tau} + \Delta^{\mathbf{N}}_{j\tau}. \tag{A.3}$$

Remark 6 *In the main text of this thesis the number of individuals diagnosed with this infection is referred as the number of observed cases and denoted as* $C_t^{obs(2)}$.

There is no epidemic at first, just some noisy observations. Then the epidemic starts in the first population at random time

$$\tau_{inf} \stackrel{\triangle}{=} \inf\{t : I^{(1)}_t > \bar{I}\},$$

where $\bar{I}$ is the epidemic threshold. After time τ_{inf} the first population infects the second

population and the epidemic starts in the second population at time

$$\tau_{lag}^{start} \triangleq \inf\{t : I_t^{(2)} > \bar{I}\}.$$

Thus, the infection goes through most of the populations in the model.

Suppose we start observing the number of infecteds in two populations at time 0. We assume that the epidemic has some probability $1/\mu_{\text{start}}$ to start in the first population in a time period $(j\tau, (j+1)\tau),\ j = 0, 1, 2, 3, \ldots$. Thus, the start of the epidemic in the first population follows the geometric distribution with mean μ_{start}/τ.

See Figure A.1 for a sample behavior of X_t and Y_t simulated using a 2 population SIR model with parameters: $\mu_{SI}^{(1)}(t) = \mu_{SI}^{(2)}(t) = 0.75$, $\gamma^{(1,2)} = \gamma^{(2,1)} = \frac{1}{20}$, $\mu_{IR}^{(1)}(t) = \mu_{IR}^{(2)}(t) = 0.5$, $\tau = 0.1$, $\tau_{inf} = 5$, $I_{\tau_{inf}}^{(1)} = 5$, $I_{\tau_{inf}}^{(2)} = 0$, $M_1 = M_2 = 2000$, $\lambda_0 = 1$, $p = 0.2$.

The first plot in the Figure A.1 shows the behavior of $\mathbf{I}_t = \{I_t^{(1)}, I_t^{(2)}\}$. We can clearly see that there were only 20 infected individuals in the first population, when the infections started in the second population.

The second plot in the Figure A.1 shows the cumulative number of new infecteds by time t, i.e. $\sum_{s=0}^{t} \Delta_s^I$. We can see that approximately 60 people has been infected by the time the epidemic started in the second population.

The last plot shows the behavior of $\mathbf{Y}_{j\tau},\ j = 0, 1, 2, 3, \ldots$. Thus, we observe that before time τ_{inf} the number of infecteds is close to the expected value of the noise $E(N_{j\tau})$. However after time τ_{inf} and t_{lag}^{start} the number of infected individuals in the respected populations fluctuates away from $E(N_{j\tau})$. Since process $\mathbf{Y}$ includes the noisy observations the cumulative numbers of infected individuals in the first and second populations by time τ_{inf} and t_{lag}^{start}, respectively, is greater than the ones on the second plot. We also see that the rate at which the cumulative number of infected individuals grows slower on the third plot than on the second plot since the number of noisy observations is smaller than the

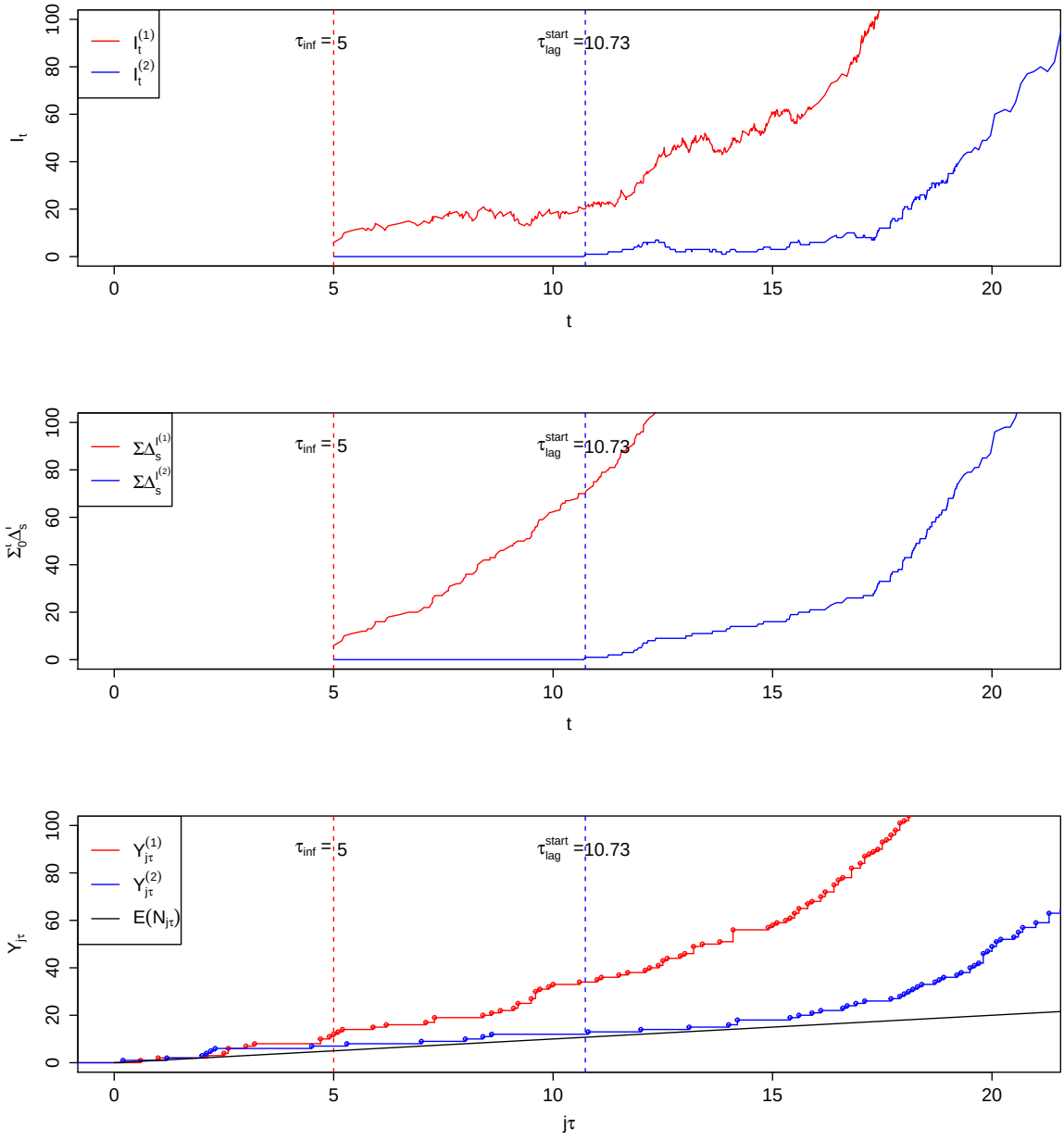


Figure A.1: First plot: Actual number of Infecteds in two populations at a given time. Second plot: cumulated number of Infecteds in two populations. Third plot: Observed number of Infecteds in two populations.

number of unobserved infections.

A.2 Inference of $\mathbf{X}_{[0,J]}$ Given $\mathbf{Y}_{[0,J]}$

We assume that we observe $\mathbf{Y}_{[0,J]} = \{\mathbf{Y}_{j\tau}, j = 0, 1, 2, \ldots, J\}$ as well as all parameters of SIR model, μ_{start}, λ_0 and p. Let Θ to be the set of all known parameters, i.e. $\Theta = \{\mu_{SI}^{(1)}(t), \mu_{SI}^{(2)}(t), \mu_{IR}^{(1)}(t), \mu_{IR}^{(2)}(t), M, \lambda_0, p, \mu_{\text{start}}\}$.

By Bayes rule we can compute the posterior distribution of $\mathbf{X}_{[0,J]} = \{\mathbf{X}_{j\tau}, j = 0, 1, 2, \ldots, J\}$ as

$$p_{\mathbf{X}_J|\mathbf{Y}_{[0,J]}}\left(\mathbf{x}_J|\mathbf{y}_{[0,J]}\right) = \frac{p_{\mathbf{Y}_J|\mathbf{X}_J,\mathbf{Y}_{[0,J-1]}}\left(\mathbf{y}_J|\mathbf{x}_J,\mathbf{y}_{[0,J-1]}\right)\pi_{\mathbf{X}_J|\mathbf{Y}_{[0,J-1]}}\left(\mathbf{x}_J|\mathbf{y}_{[0,J-1]}\right)}{\int p_{\mathbf{Y}_J|\mathbf{X}_J,\mathbf{Y}_{[0,J-1]}}\left(\mathbf{y}_J|\mathbf{x}_J,\mathbf{y}_{[0,J-1]}\right)\pi_{\mathbf{X}_J|\mathbf{Y}_{[0,J-1]}}\left(\mathbf{x}_J|\mathbf{y}_{[0,J-1]}\right)d\mathbf{x}_J}, \tag{A.4}$$

where X_j has initial density function π_{X_0} and transition probabilities $\pi_{X_j|X_{(j-1)}}(x'|x)$ such that

$$\pi_{X_{[0,J]}} = \pi_{X_0}\prod_{j=1}^{n}\pi_{X_j|X_{j-1}}(x_j|x_{j-1}).$$

However since the denominator in (A.4) is just a normalizing constant, we get

$$p_{\mathbf{X}_J|\mathbf{Y}_{[0,J]}}\left(\mathbf{x}_J|\mathbf{y}_{[0,J]}\right) \propto p_{\mathbf{Y}_J|\mathbf{X}_J,\mathbf{Y}_{[0,J-1]}}\left(\mathbf{y}_J|\mathbf{x}_J,\mathbf{y}_{[0,J-1]}\right)\pi_{\mathbf{X}_J|\mathbf{Y}_{[0,J-1]}}\left(\mathbf{x}_J|\mathbf{y}_{[0,J-1]}\right). \tag{A.5}$$

Then if we condition on the path $X_{(J-1,J)}$, we see (A.5) is proportional to

$$\int p_{\mathbf{Y}_J|\mathbf{X}_{(J-1,J]},\mathbf{Y}_{[0,J-1]}}\left(\mathbf{y}_J|\mathbf{x}_{(J-1,J]},\mathbf{y}_{[0,J-1]}\right)\pi_{\mathbf{X}_{(J-1,J]}|\mathbf{Y}_{[0,J-1]}}\left(\mathbf{x}_{(J-1,J]}|\mathbf{y}_{[0,J-1]}\right)d\mathbf{x}_{(J-1,J)}. \tag{A.6}$$

First, let's look at the first term of the right hand side of the equation (A.6). By

construction,

$$p_{\mathbf{Y}_J|\mathbf{X}_{(J-1,J]},\mathbf{Y}_{[0,J-1]}}\left(\mathbf{y}_J|\mathbf{x}_{(J-1,J]},\mathbf{y}_{[0,J-1]}\right)=\prod_{k=1}^{K}p_{\mathbf{Y}_J^{(k)}|\mathbf{X}_{(J-1,J]}^{(k)},\mathbf{Y}_{[0,J-1]}^{(k)}}\left(\mathbf{y}_J^{(k)}|\mathbf{x}_{(J-1,J]}^{(k)},\mathbf{y}_{[0,J-1]}^{(k)}\right). \tag{A.7}$$

Observe that if we know the paths of $\mathbf{Y}_{[0,J]}^{(k)}$ and $\mathbf{X}_{(J-1,J]}^{(k)}$, then we know $\Delta\mathbf{Y}_J^{(k)}$ and $\Delta\mathbf{I}_J^{(k)}$. Thus, we observe that

$$p_{\mathbf{Y}_J^{(k)}|\mathbf{X}_{(J-1,J]}^{(k)},\mathbf{Y}_{[0,J-1]}^{(k)}}\left(\mathbf{y}_J^{(k)}|\mathbf{x}_{(J-1,J]}^{(k)},\mathbf{y}_{[0,J-1]}^{(k)}\right)=p_{\Delta\mathbf{Y}_J^{(k)}|\Delta\mathbf{I}_J^{(k)}}\left(\Delta\mathbf{y}^{(k)}|\Delta\mathbf{i}^{(k)}\right), \tag{A.8}$$

where $\Delta\mathbf{y}^{(k)}$ and $\Delta\mathbf{i}^{(k)}$ are vector of values, which processes $\Delta\mathbf{Y}_J^{(k)}$ and $\Delta\mathbf{I}_J^{(k)}$ take, respectively. Then condition on $\Delta\mathbf{N}_J^{(k)}$ we get

$$p_{\Delta\mathbf{Y}_J^{(k)}|\Delta\mathbf{I}_J^{(k)}}\left(\Delta\mathbf{y}^{(k)}|\Delta\mathbf{i}^{(k)}\right)=\sum_{l=1}^{\infty}p_{\Delta\mathbf{Y}_J^{(k)}|\Delta\mathbf{I}_J^{(k)},\Delta\mathbf{N}_J^{(k)}}\left(\Delta\mathbf{y}^{(k)}|\Delta\mathbf{i}^{(k)},l\right)P(\Delta\mathbf{N}_J^{(k)}=l), \tag{A.9}$$

where we substitute the corresponding density functions from distributions (A.1) and (A.2).

Thus, substituting (A.8), (A.9) into (A.7), we get

$$\begin{aligned}&p_{\mathbf{Y}_J|\mathbf{X}_{(J-1,J]},\mathbf{Y}_{[0,J-1]}}\left(\mathbf{y}_J|\mathbf{x}_{(J-1,J]},\mathbf{y}_{[0,J-1]}\right)\\&\qquad=\prod_{k=1}^{K}\left[\sum_{l=1}^{\infty}p_{\Delta\mathbf{Y}_J^{(k)}|\Delta\mathbf{I}_J^{(k)},\Delta\mathbf{N}_J^{(k)}}\left(\Delta\mathbf{y}^{(k)}|\Delta\mathbf{i}^{(k)},l\right)P(\Delta\mathbf{N}_J^{(k)}=l)\right].\end{aligned} \tag{A.10}$$

The second term of the right hand side of (A.6)

$$\pi_{\mathbf{X}_{(J-1,J]}|\mathbf{Y}_{[0,J-1]}}\left(\mathbf{x}_{(J-1,J]}|\mathbf{y}_{[0,J-1]}\right)$$

can be rewritten conditioning on the state $\mathbf{X}_{j-1}$ as well as applying Markov property of

$\mathbf{X}_j$ as

$$\int \pi_{X_{[J-1,J]}}\left(\mathbf{x}_{[J-1,J]}\right) \pi_{X_{[J-1,J]}|Y_{[0,J-1]}}\left(\mathbf{x}_{[J-1,J]}|\mathbf{y}_{[0,J-1]}\right) d\mathbf{x}_{J-1}. \tag{A.11}$$

As the size of populations M_j increases, $\pi_{X_{[J-1,J]}}\left(\mathbf{x}_{(J-1,J]}\right)$ is deterministic. Therefore, we get that the posterior likelihood of $\mathbf{X}_J|\mathbf{Y}_{[0,J]}$ is proportional to

$$\iint [\sum_{l=1}^{\infty} p_{\Delta\mathbf{Y}_J^{(k)}|\Delta\mathbf{I}_J^{(k)},\Delta\mathbf{N}_J^{(k)}}\left(\Delta\mathbf{y}^{(k)}|\Delta\mathbf{i}^{(k)},l\right) P(\Delta\mathbf{N}_J^{(k)}=l)]$$
$$\pi_{X_{[J-1,J]}|Y_{[0,J-1]}}\left(\mathbf{x}_{[J-1,J]}|\mathbf{y}_{[0,J-1]}\right) d\mathbf{x}_{J-1}d\mathbf{x}_{(J-1,J)}. \tag{A.12}$$

We use sequential methods since the integral in (A.12) depends on the previous step, i.e. calculation of

$$\pi_{X_{[J-1,J]}|Y_{[0,J-1]}}\left(\mathbf{x}_{[J-1,J]}|\mathbf{y}_{[0,J-1]}\right).$$

Thus, the integral in (A.12) is not possible to compute analytically. Therefore we implement the Sequential Monte Carlo method (SMC), or particle filtering.

A.3 Particle Filtering

SMC methods is a combination of Sequential Importance Sampling (SIS) and Resampling steps. Suppose we have n particles $i = 1,\ldots,n$. Each particle behaves independently and its behavior is defined by $\mathbf{X}_t^i = \{\mathbf{S}_t^i, \mathbf{I}_t^i\}$ for $i = 1,\ldots,n$, where $\mathbf{S}_t^i = \{S_t^{(1)^i},\ldots,S_t^{(K)^i}\}$ and $\mathbf{I}_t^i = \{I_t^{(1)^i},\ldots,I_t^{(K)^i}\}$. Then there are three cases of particle behavior:

- If an ith particle's epidemic starts at time $j\tau$, then the particle gets positioned at $\mathbf{X}_{j\tau}^i = \mathbf{X}_0$, which was sampled from the distribution $\pi(\mathbf{X}_0)$.

- If an ith particle's epidemic started before time $j\tau$, then we simulate the behavior of SIR model from its previous position $\mathbf{X}^i_{(j-1)\tau}$ using the tau-leaping algorithm in time interval $((j-1)\tau, j\tau)$ with the reactions channels along with the rate functions for each population defined in (3.1) and the reaction channels along with the rate functions for transmission defined in (3.28).

- If an ith particle's epidemic didn't start before time $j\tau$, then it gets positioned at $\mathbf{X}^i_{j\tau} = \{\mathbf{M}, \mathbf{0}\}$.

Then for all particles we define the number of new infections for particle i in a time interval $((j-1)\tau, j\tau)$ to be

$$\Delta^{I^i}_{j\tau} = \mathbf{S}^i_{(j-1)\tau} - \mathbf{S}^i_{j\tau} \tag{A.13}$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots$. Note that $\Delta\mathbf{I}^i_{j\tau} = \{\Delta\mathbf{I}^{(1)^i}_{j\tau}, \ldots, \Delta\mathbf{I}^{(K)^i}_{j\tau}\}$, where $\Delta\mathbf{I}^{(k)^i}_{j\tau}$ is the number of new infections for particle i in the population $k = 1, \ldots, K$.

First, at time $j = 1$ we sample $\mathbf{X}^i_{j\tau}$ from the SIR model for each particle i given the starting values $\mathbf{X}_{j-1}$. Then we compute its weight $w_i(j\tau)$ as

$$w_i(j\tau) = w_i((j-1)\tau) \times p_{\Delta\mathbf{Y}_{J\tau}|\Delta\mathbf{I}_{J\tau}}\left(\Delta\mathbf{y}|\Delta\mathbf{i}\right), \tag{A.14}$$

where $w_i(0) = 1\ \forall i$. We can rewrite formula (A.14) using formulas (A.1), (A.2) (A.9), and (A.13) as:

$$w_i(j\tau) = w_i((j-1)\tau) \times \prod_{k=1}^{K} \left[\sum_{l=0}^{\infty} \binom{\Delta\mathbf{Y}_{j\tau}}{\Delta\mathbf{I}^{(k)^i}_{j\tau} - l} p^{\Delta\mathbf{I}^{(k)^i}_{j\tau}} (1-p)^{\Delta\mathbf{I}^{(k)^i}_{j\tau} - l} \cdot e^{-\lambda_0\tau} \frac{(\lambda_0\tau)^l}{l!} \right], \tag{A.15}$$

$j = 1, 2, \ldots, J$, where $w_i(0) = 1\ \forall i = 1, \ldots, n$.

Second, we use a resampling step to remove the particles with low weights and increase the number of particles with high weights. We use the Effective Sample Size (ESS)

criterion to evaluate if a residual resampling is needed Doucet and Johansen [2009]. If ESS criterion is satisfied, we resample using weights $w_i(j\tau)$ to obtain n equally weighted particles $\bar{\mathbf{X}}_t^i$.

Thus, we repeat the steps above and we get that the posterior probability measure of $\mathbf{X}_{[0,J\tau]}|\mathbf{Y}_{[0,J\tau]}$ is equal to

$$\mathcal{L}_{\mathbf{X}_{J\tau}|\mathbf{Y}_{[0,J\tau]}}(\mathbf{x}|\mathbf{y}) \approx \frac{1}{n}\sum_{i=1}^{n} w_i(j\tau)\delta_{\mathbf{X}_{j\tau}^i}(\mathbf{x}), \quad j = 1, 2, \ldots, J,$$

where $\delta_{x_0}(x)$ denotes the Dirac delta mass located at x_0.

The algorithm of the above estimation is given in Algorithm B.7. Figure A.2 provides 95% confidence region of $\mathbf{X}_t$ for a two population model simulated under parameters: $n = 1000$, $\mu_{start} = 30$, $p = 0.2$, $\lambda_0 = 1$, $\tau = 0.1$, $\mu_{SI}^{(1)}(t) = \mu_{SI}^{(2)}(t) = 0.75$, $\mu_{SI}^{(1,2)}(t) = 0.0375$, $\mu_{IR}^{(1)}(t) = 0.5$, $M_1 = M_2 = 2000$. Before time τ_{inf}, $I_{j\tau}^i$ has stationary distribution: some

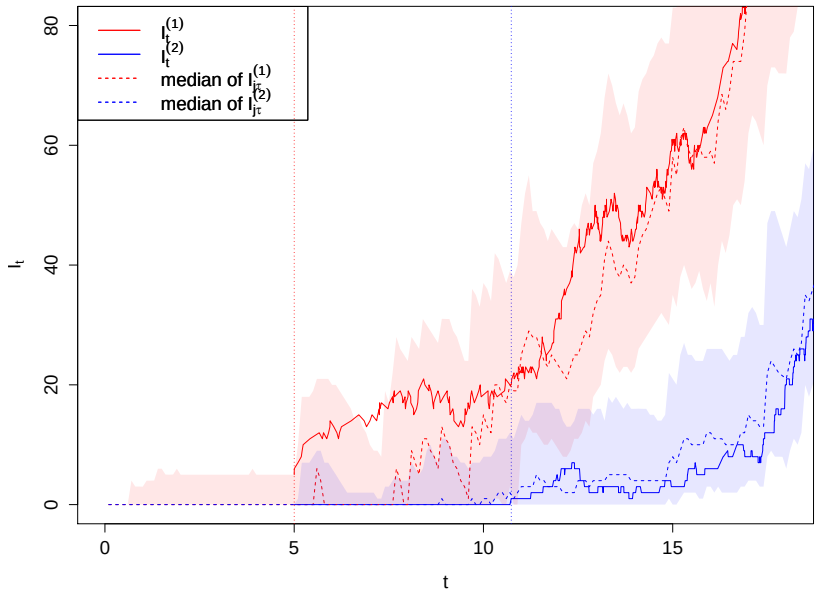


Figure A.2: Estimated 95% quantile interval of $\mathbf{X}_t$ from $\mathbf{Y}_t$

particles think that there is some infection going on in the first population, however it is not crucial enough. After time τ_{inf} number of infected in the first population increases, which causes the a start of epidemic in the second population. We see that the 95% quantile regions cover almost the whole $\mathbf{X}_t$.

As we increase λ_0 or decrease p, the quantile regions become wider since there is more uncertainty in the model. A decrease in μ_{start} below 20 in our case, i.e. an increase in the probability of epidemic to start, destroys the stationarity before time τ_{inf} since more particles think that epidemic started.

We denote the estimate posterior probability that the epidemic started as

$$P(\mathbf{I}^i_{j\tau} > 0 \ \forall i = 1, \ldots, n | \mathbf{Y}_{[0,J\tau]})$$

for $j = 0, 1, \ldots, J$, and we compute it as the sum of the weights of particles for which an epidemic started. The true value of the probability that epidemic started is either 0 ("no epidemic") or 1("epidemic"). However the estimate of this probability can take any value between 0 and 1.

The graph of the posterior probability is given on Figure A.3. Note that this plot corresponds to the simulation given on Figure A.2. We see that there is a significant increase in the likelihood that epidemic started in the first population right after the actual start of epidemic. The start of the first epidemic causes an increase of the probability of the second epidemic started. As discussed before we might choose a level of the probability for which we will treat the epidemic as started. If in this example we choose it to be 0.5, then we see that we would consider the epidemic started in the second population as "started" too early. Thus, there is a false-alarm. If we choose the benchmark to be 0.8, then we detected the epidemic late.

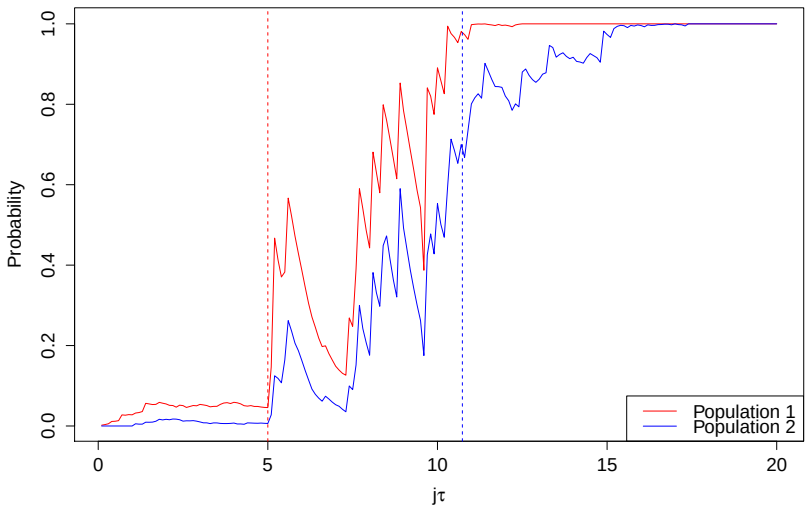


Figure A.3: $P(\mathbf{I}^i_{j\tau} > 0 \ \forall i | \mathcal{G}_{j\tau})$ vs. time, where the dashed lines represent the actual start time of epidemic

Appendix B

Algorithms

Algorithm B.1 Path Generation of First Pseudo-Posterior

Require: σ_δ^2, $\mu_{SI}^{(1,2)}(t)$, $\mathbf{X}_{0:T}^{(1)}$, T

1: **for** $t = 0, ..., T$ **do**
2: Compute P_{t+1} via Equation 4.3
3: **end for**
4: **return** $(P_t)\,,\ t = 0, 1, 2, \ldots, T$

Algorithm B.2 Path Generation of Second Pseudo-Posterior via 'Original' procedure

Require: σ_0^2, $\mathbf{X}_{0:T}^{(1)}$, $C_{0:T}^{obs(2)}$, T, $\eta_{0:T}^{(2)}$

1: **for** $t = 0, ..., T$ **do**
2: Compute μ_{t+1} and σ_{t+1} via Theorem (2)
3: **end for**
4: **return** $\left(\mathbf{X}_t^{(1)}, \mu_t, \sigma_t^2\right),\ t = 0, 1, 2, \ldots, T$

Algorithm B.3 Path Generation of Second Pseudo-Posterior via 'Gamma-Poisson' procedure

Require: μ_0, σ_0^2, $\mathbf{X}_{0:T}^{(1)}$, T, $\eta_{0:T}^{(2)}$

Compute α_0 and β_0 via Equations (4.30)
for $t = 0, ..., T$ **do**
    Sample $I_t^{(2)}$ from $Gamma(\alpha_t, \beta_t)$
    Simulate $C_{t+1}^{obs(2)} \sim Poisson\left((1-\rho)\beta_1 \eta_t^{(2)} I_t^{(2)}\right)$
    Compute μ_{t+1} and σ_{t+1} via Theorem (2)
    Compute α_{t+1} and β_{t+1} via Equations (4.30)
end for
return $\left(\mathbf{X}_t^{(1)}, \mu_t, \sigma_t^2\right),\ t = 0, 1, 2, \ldots, T$

Algorithm B.4 Path Generation of Second Pseudo-Posterior via 'Negative Binomial' procedure

Require: μ_0, σ_0^2, $\mathbf{X}_{0:T}^{(1)}$, T, $\eta_{0:T}^{(2)}$

Compute α_0 and β_0 via Equations (4.30)
for $t = 0, ..., T$ **do**
    Simulate $C_{t+1}^{obs(2)} \sim NegBin\left(\alpha_t, \beta_t/((1-\rho)\beta_1 \eta_t^{(2)} + \beta_t)\right)$
    Compute μ_{t+1} and σ_{t+1} via Theorem (2)
    Compute α_{t+1} and β_{t+1} via Equations (4.30)
end for
return $\left(\mathbf{X}_t^{(1)}, \mu_t, \sigma_t^2\right),\ t = 0, 1, 2, \ldots, T$

Algorithm B.5 Path and Cost Generation

Require: $\{\mathbf{x}_0^n\}_{n=1}^N$, $\mathfrak{S}_{0:t-1}$

for $n = 1, \ldots, N$ **do**
    $s \leftarrow 1$
    while $s \leq t$ **do**
        Simulate the next state $\mathbf{x}_s^n \sim p_1(\cdot|\mathbf{x}_{s-1}^n)$
        if $\mathbf{x}_s^n \in \mathfrak{S}_{t-s}$ **then** Break
        end if
        $s \leftarrow s + 1$
    end while
    $\tau_t^n \leftarrow s$
    Compute $q^n \equiv c(\mathbf{x}_{0:\tau_t^n}^n) - d(\mathbf{x}_0^n)$ using formula (2.2) and (2.3)
end for
return $\{(\mathbf{x}_0^n, q^n)\}_{n=1}^N$

Algorithm B.6 Sequential Regression Monte Carlo

Require: $C_{\text{FA}}, C_{\text{Delay}}$, N_0, N', N^{end}, D, Regression Model

$\hat{\mathfrak{S}}_0 \leftarrow \mathcal{X}$
for $t = 1, 2, \ldots$ **do**
    Generate experimental design $\{\mathbf{x}_0^n,\ n = 1, \ldots, N_0\}$
    Compute the difference of scenario future costs and immediate costs $q^n = c(\mathbf{x}_{0:\tau_t^n}^n) - d(\mathbf{x}_0^n)$ for $n = 1, \ldots, N_0$ using Algorithm B.5 and $\hat{\mathfrak{S}}_{0:t-1}$
    $Z \leftarrow \{(\mathbf{x}_0^n, q^n)\}_{n=1}^{N_0}$
    Regress $\{q^n\}$ on $\{\mathbf{x}_0^n\},\ n = 1, \ldots, N_0$ using Loess (5.7)
    Initialize $N \leftarrow N_0$
    while $N < N^{end}$ **do**
        Generate X_{finite} of size D using Latin Hypercube Sampling on $\mathcal{X}$
        Compute the acquisition weights $w(\mathbf{x})\ \forall \mathbf{x} \in X_{\text{finite}}$ via (5.11) and (5.9)
        Sample $\{\mathbf{x}_0^n\}_{n=N+1}^{N+N'}$ from X_{finite} using weights $w(\mathbf{x})$
        Simulate the costs q^n, $n = N+1, \ldots, N+N'$ using Algorithm B.5
        $Z \leftarrow Z \cup \{(\mathbf{x}_0^n, q^n)\}_{n=N+1}^{N+N'}$
        Update the Regression model (5.7) using the latest Z
        $N \leftarrow N + N'$
    end while
    $\hat{\mathfrak{S}}_t \leftarrow \{\mathbf{x} \in \mathcal{X} : \hat{q}(t, \mathbf{x}) > 0\}$, cf. (5.4)
end for

Algorithm B.7 Estimation of $\mathbf{X}_t$ from $\mathbf{Y}_t$ in a multiple populations case

Sample $\mathbf{X}_0$
for $j \leftarrow 1,J$ **do**
    Observe $\Delta^Y_{j\tau}$
    for $i \leftarrow 1, n$ **do**
        if epidemic of particle i started at step j **then**
            Particle position is at $\mathbf{X}^i_{j\tau} = \mathbf{X}_0$
        else if epidemic of particle i started before step j **then**
            Simulate a particle behavior $\mathbf{X}^i_{j\tau}$ from $\mathbf{X}^i_{(j-1)\tau}$
        else
            Particle position is at $\mathbf{X}^i_{j\tau} = \{\mathbf{M}, \mathbf{0}\}$
        end if
        Calculate $\Delta^{I^i}_{j\tau}$ using formula (A.13)
        Calculate $w_i(j\tau)$ using formula (A.15)
    end for
    if resampling criterion is satisfied **then**
        Resample $\{w_i(j\tau), \mathbf{X}^i_{j\tau}\}$ to obtain n equally weighted particles $\bar{\mathbf{X}}^i_{j\tau}$
        Set $\{w_i(j\tau), \mathbf{X}^i_{j\tau}\} \leftarrow \{\frac{1}{n}, \bar{\mathbf{X}}^i_{j\tau}\}$
    end if
end for
return $\mathcal{L}_{\mathbf{X}_{J\tau}|\mathbf{Y}_{[0,J\tau]}}(\mathbf{x}|\mathbf{y})$

Bibliography

Ebola Virus Disease in West Africa: The first 9 months of the epidemic and forward projections. *New England Journal of Medicine*, 371(16):1481–1495, 2014. PMID: 25244186.

M. Ajelli, B. Gonçalves, D. Balcan, V. Colizza, H. Hu, J. Ramasco, S. Merler, and A. Vespignani. Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models. *BMC Infectious Diseases*, 10 (1):190, 2010.

Linda J.S. Allen and Amy M. Burgin. Comparison of deterministic and stochastic SIS and SIR models in discrete time. *Mathematical Biosciences*, 163(1):1 – 33, 2000. ISSN 0025-5564.

Linda JS Allen, Fred Brauer, Pauline Van den Driessche, and Jianhong Wu. *Mathematical epidemiology*. Springer, 2008.

Linda JS Allen, Yuan Lou, and Andrew L Nevai. Spatial patterns in a discrete-time SIS patch model. *Journal of Mathematical Biology*, 58(3):339–375, 2009.

H. Andersson and T. Britton. *Stochastic Epidemic Models and Their Statistical Analysis*. Lecture Notes in Statistics Series. Springer Verlag, 2000. ISBN 9780387950501.

Frank Ball and Damian Clancy. The final size and severity of a generalised stochastic multitype epidemic model. *Advances in Applied Probability*, 25(4):721–736, 1993.

Frank Ball and Peter Donnelly. Strong approximations for epidemic models. *Stochastic Processes and their Applications*, 55(1):1–21, 1995.

David Banks, Gauri Datta, Alan Karr, James Lynch, Jarad Niemi, and Francisco Vera. Bayesian CAR models for syndromic surveillance on multiple data streams: Theory and practice. *Information Fusion*, 13:105–116, 2012. ISSN 1566-2535.

D. P. Bertsekas. *Dynamic programming and optimal control. Vol. I.* Athena Scientific, Belmont, MA, third edition, 2005. ISBN 1-886529-26-4.

Ottar N. Bjørnstad, Bärbel F. Finkenstädt, and Bryan T. Grenfell. Dynamics of measles epidemics: Estimating scaling of transmission rates using a time series sir model. *Ecological Monographs*, 72(2):169–184, 2002. ISSN 1557-7015. doi: 10.1890/0012-9615(2002)072[0169:DOMEES]2.0.CO;2. URL `http://dx.doi.org/10.1890/0012-9615(2002)072[0169:DOMEES]2.0.CO;2`.

Fred Brauer. Compartmental models in epidemiology. In *Mathematical epidemiology*, pages 19–79. Springer Berlin Heidelberg, 2008.

Ta-Chien Chan, Chwan-Chuen King, Muh-Yong Yen, Po-Huang Chiang, Chao-Sheng Huang, and Chuhsing K Hsiao. Probabilistic daily ILI syndromic surveillance with a spatio-temporal Bayesian hierarchical model. *PLoS One*, 5(7):e11626, 2010.

Gerardo Chowell and Hiroshi Nishiura. Transmission dynamics and control of Ebola virus disease (EVD): a review. *BMC Medicine*, 12(1):196, 2014.

William S. Cleveland and Susan J. Devlin. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83 (403):pp. 596–610, 1988. ISSN 01621459.

Andrew JK Conlan and Bryan T Grenfell. Seasonality and the persistence and invasion of measles. *Proceedings of the Royal Society of London B: Biological Sciences*, 274 (1614):1133–1141, 2007.

Benjamin J. Cowling, Irene O. L. Wong, Lai-Ming Ho, Steven Riley, and Gabriel M. Leung. Methods for monitoring influenza surveillance data. *International Journal of Epidemiology*, 35:1314–1321, 2006.

Aron Culotta. Towards detecting influenza epidemics by analyzing Twitter messages. In *Proceedings of the first workshop on social media analytics*, pages 115–122. ACM, 2010.

Arnaud Doucet and Adam M. Johansen. A tutorial on particle filtering and smoothing: fifteen years later. *Handbook of nonlinear filtering*, 12(3):656–704, 2009.

V. Dukic, H.F. Lopes, and N.G. Polson. Tracking epidemics with Google Flu Trends data and a state-space SEIR model. *Journal of the American Statistical Association*, 107(500):1410–1426, 2012.

D. Egloff. Monte Carlo algorithms for optimal stopping and statistical learning. *Annals of Applied Probability*, 15(2):1396–1432, 2005.

C. P. Farrington, N. J. Andrews, A. D. Beale, and M. A. Catchpole. A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 159(3):547–563, 1996. ISSN 09641998, 1467985X. URL `http://www.jstor.org/stable/2983331`.

David Fisman, Edwin Khoo, and Ashleigh Tuite. Early epidemic dynamics of the West African 2014 Ebola outbreak: estimates derived with a simple two-parameter model. *PLoS Currents Outbreaks*, 6:Sep 8, 2014.

B. Gramacy and M. Ludkovski. Sequential Design for Optimal Stopping Problems. *SIAM Journal on Financial Mathematics*, 6(1):748–775, 2015.

Nicholas C. Grassly, Christophe Fraser, and Geoffrey P. Garnett. Host immunity and synchronized epidemics of syphilis across the united states. *Nature*, 433(7024):417–421, 01 2005. URL `http://dx.doi.org/10.1038/nature03072`.

Jean-Olivier Guintran, Charles Delacollette, Peter I Trigg, Malaria Unit, World Health Organization, et al. Systems for the early detection of malaria epidemics in africa: an analysis of current practices and future priorities. 2006.

T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference and prediction*. Springer, 2009.

D. He, E.L. Ionides, and A.A. King. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *Journal of the Royal Society Interface*, 7(43):271–283, 2010.

HW Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, DEC 2000.

Ruimeng Hu and Mike Ludkovski. Sequential design for ranking response surfaces. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):212–239, 2017. doi: 10.1137/15M1045168. URL `http://dx.doi.org/10.1137/15M1045168`.

LC Hutwagner, WW Thompson, GM Seeman, and T Treadwell. A simulation model for assessing aberration detection methods used in public health surveillance for systems with limited baselines. *Statistics in Medicine*, 24(4):543–550, 2005.

G. James, T. Hastie, D. Witten, and R. Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics. Springer London, Limited, 2013. ISBN 9781461471370.

Matt J Keeling and Ken TD Eames. Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4):295–307, 2005.

Matt J Keeling and Bryan T Grenfell. Understanding the persistence of measles: reconciling theory, simulation and observation. *Proceedings of the Royal Society of London B: Biological Sciences*, 269(1489):335–343, 2002.

M.J. Keeling and P. Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2011. ISBN 9781400841035. URL `https://books.google.com/books?id=LxzILSuKDhUC`.

Martin Kulldorff, Richard Heffernan, Jessica Hartman, Renato Assuncao, and Farzad Mostashari. A space–time permutation scan statistic for disease outbreak detection. *PLoS Med*, 2(3):e59, 02 2005.

Andrew B Lawson. *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. CRC Press, 2013.

Y. LeStrat and F. Carrat. Monitoring epidemiologic surveillance data using hidden Markov models. *Statistics in Medicine*, 18:3463–3478, 1999.

Junjing Lin and Michael Ludkovski. Sequential Bayesian inference in Hidden Markov stochastic kinetic models with application to detection and response to seasonal epidemics. *Statistics and Computing*, 24(6):1047–1062, 2014.

Wei-min Liu, Simon A Levin, and Yoh Iwasa. Influence of nonlinear incidence rates upon the behavior of sirs epidemiological models. *Journal of mathematical biology*, 23(2): 187–204, 1986.

M. Ludkovski. A simulation approach to optimal stopping under partial observations. *Stochastic Processes and Applications*, 119(12):4061–4087, 2009.

M. Ludkovski and J. Niemi. Optimal dynamic policies for influenza management. *Statistical Communications in Infectious Diseases*, 2(1):article 5 (electronic), 2010.

Michael Ludkovski. Monte Carlo methods for adaptive disorder problems. In Rene A. Carmona, Pierre Del Moral, Peng Hu, and Nadia Oudjane, editors, *Numerical Methods in Finance*, volume 12 of *Springer Proceedings in Mathematics*, pages 83–112. Springer Berlin Heidelberg, 2012.

Mike Ludkovski and Jarad Niemi. Optimal disease outbreak decisions using stochastic simulation. In *Proceedings of the Winter Simulation Conference*, WSC '11, pages 3849–3858. Winter Simulation Conference, 2011.

D.J.C. MacKay. Information-based objective functions for active data selection. *Neural computation*, 4(4):590–604, 1992.

MA Martínez-Beneito, David Conesa, Antonio López-Quílez, and Aurora López-Maside. Bayesian Markov switching models for the early detection of influenza epidemics. *Statistics in Medicine*, 27(22):4455–4468, 2008.

D.C. Montgomery, E.A. Peck, and G.G. Vining. *Introduction to Linear Regression Analysis*. Wiley Series in Probability and Statistics. Wiley, 2015. ISBN 9781119180173. URL `https://books.google.com/books?id=27kOCgAAQBAJ`.

Peter Neal. The basic reproduction number and the probability of extinction for a dynamic epidemic model. *Mathematical Biosciences*, 236(1):31–35, 2012.

D.B. Neill. Fast Bayesian scan statistics for multivariate event detection and visualization. *Statistics in Medicine*, 30(5):455–469, 2011.

Vesna Nevistic and James A. Primbs. Model predictive control: Breaking through constraints. In *35th IEEE Conference on Decision and Control*, pages 3932–3937, 1996.

Hiroshi Nishiura. Real-time forecasting of an epidemic using a discrete time stochastic model: a case study of pandemic influenza (h1n1-2009). *BioMedical Engineering OnLine*, 10(1):15, 2011. ISSN 1475-925X.

Angela Noufaily, Doyo G Enki, Paddy Farrington, Paul Garthwaite, Nick Andrews, and Andre Charlett. An improved algorithm for outbreak detection in multiple surveillance systems. *Statistics in Medicine*, 32(7):1206–1222, 2013.

H Vincent Poor and Olympia Hadjiliadis. *Quickest detection*, volume 40. Cambridge University Press, 2009.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.

Leonid A Rvachev and Ira M Longini. A mathematical model for the global spread of influenza. *Mathematical Biosciences*, 75(1):3–22, 1985.

Dieter Schenzle. An age-structured model of pre- and post-vaccination measles transmission. *Mathematical Medicine and Biology*, 1(2):169–191, 1984. doi: 10.1093/imammb/1.2.169. URL `http://imammb.oxfordjournals.org/content/1/2/169.abstract`.

Paola Sebastiani, Kenneth D Mandl, Peter Szolovits, Isaac S Kohane, and Marco F Ramoni. A Bayesian dynamic model for influenza surveillance. *Statistics in Medicine*, 25(11):1803–1816, 2006.

Ekaterina Shatskikh and Michael Ludkovski. Computational method for epidemic detection in multiple populations. *Online Journal of Public Health Informatics*, 7(1): e158, 2015. doi: 10.5210/ojphi.v7i1.5824. URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4512471/`.

K. Shatskikh and M. Ludkovski. Bayesian Epidemic Detection in Multiple Populations. *ArXiv e-prints*, September 2015.

K. Shatskikh and M. Ludkovski. Bayesian inference and detection for a multi-population SIR model of stochastic epidemics, 2017. In preparation.

Daniel M Sheinson, Jarad Niemi, and Wendy Meiring. Comparison of the performance of particle filter algorithms applied to tracking of a disease epidemic. *Mathematical Biosciences*, 255:21–32, 2014.

Galit Shmueli and Howard Burkom. Statistical challenges facing early outbreak detection in biosurveillance. *Technometrics*, 52(1):39–51, 2010.

Alex Skvortsov and Branko Ristic. Monitoring and prediction of an epidemic outbreak using syndromic observations. *Mathematical Biosciences*, 240(1):12 – 19, 2012. ISSN 0025-5564.

Hal L Smith, Liancheng Wang, and Michael Y Li. Global dynamics of an SEIR epidemic model with vertical transmission. *SIAM Journal on Applied Mathematics*, 62(1):58–69, 2001.

Matthew W. Tanner, Lisa Sattenspiel, and Lewis Ntaimo. Finding optimal vaccination strategies under parameter uncertainty using stochastic programming. *Mathematical Biosciences*, 215(2):144–151, 2008. ISSN 0025-5564.

Helen J Wearing, Pejman Rohani, and Matt J Keeling. Appropriate models for the management of infectious diseases. *PLoS Medicine*, 2(7):e174, 07 2005.

Darren J. Wilkinson. *Stochastic Modelling for Systems Biology*. Chapman & Hall/CRC, London, 2006.

Heng Yang, Olympia Hadjiliadis, and Michael Ludkovski. Quickest detection in the wiener disorder problem with post-change uncertainty. *Stochastics*, 89(3-4):654–685, 2017. doi: 10.1080/17442508.2016.1276908. URL `http://www.tandfonline.com/doi/abs/10.1080/17442508.2016.1276908`.

Wan Yang, Alicia Karspeck, and Jeffrey Shaman. Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics. *PLoS Computational Biology*, 10(4):e1003583, 2014.

Huafeng Zhou and Andrew B Lawson. EWMA smoothing and Bayesian spatial modeling for health surveillance. *Statistics in Medicine*, 27(28):5907–5928, 2008.