

A Bayesian Nonparametric Spiked Process Prior for Dynamic Model Selection

Alberto Cassese^{*†}, Weixuan Zhu^{*‡}, Michele Guindani[§], and Marina Vannucci[¶]

Abstract. In many applications, investigators monitor processes that vary in space and time, with the goal of identifying temporally persistent and spatially localized departures from a baseline or “normal” behavior. In this manuscript, we consider the monitoring of pneumonia and influenza (P&I) mortality, to detect influenza outbreaks in the continental United States, and propose a Bayesian nonparametric model selection approach to take into account the spatio-temporal dependence of outbreaks. More specifically, we introduce a zero-inflated conditionally identically distributed species sampling prior which allows borrowing information across time and to assign data to clusters associated to either a null or an alternate process. Spatial dependences are accounted for by means of a Markov random field prior, which allows to inform the selection based on inferences conducted at nearby locations. We show how the proposed modeling framework performs in an application to the P&I mortality data and in a simulation study, and compare with common threshold methods for detecting outbreaks over time, with more recent Markov switching based models, and with spike-and-slab Bayesian nonparametric priors that do not take into account spatio-temporal dependence.

Keywords: nonparametric Bayes, variable selection, Markov random field, spatio-temporal data.

1 Introduction

In disease surveillance, investigators are often concerned with monitoring a health condition, in order to understand patterns of disease progression across space and time. An outbreak may persist only briefly, since the monitored condition is eventually expected to revert to a long-term trend, or recur and endure over time. Anomalies may be confined locally to specific regions, or instead spread to large areas, possibly after some delay.

The methods routinely employed by the U.S. Center for Disease Control and Prevention (CDC) for outbreak detection are based on thresholding a specific time series based on historical data (Muscatello et al., 2008). More sophisticated methods proposed in the literature have employed hidden Markov models (HMMs) (Madigan, 2005) or more flexible Markov switching models (Martínez-Beneito et al., 2008) to capture the transition

^{*}The authors have equally contributed to the article.

[†]Department of Methodology and Statistics, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, 6229 ER, The Netherlands, alberto.cassese@maastrichtuniversity.nl

[‡]Wang Yanan Institute for Studies in Economics (WISE), Department of Statistics, School of Economics, Xiamen University, Xiamen, 361005, China, zhuweixuan@xmu.edu.cn

[§]Department of Statistics, Donald Bren School of Information and Computer Sciences, University of California, Irvine, CA 92697-1250, USA, michele.guindani@uci.edu

[¶]Department of Statistics, Duncan Hall, Rice University, Houston, TX 77005, USA, marina@rice.edu

between the epidemic and the non-epidemic phase of a disease. A few spatio-temporal models have assumed a first-order autoregressive model on the temporal dynamics of the epidemics (see, e.g. Heaton et al., 2012; Zou et al., 2012).

In this manuscript, we investigate the modeling of possibly temporary and localized changes in disease patterns as a problem of model selection over space and time. In order to identify departures of a monitored process from a baseline or “normal” behavior, we start by assuming a two-groups modeling framework (Efron, 2004, 2008): a data point, say y , is assigned either to a null model, $H_0 : y \sim F(y; \theta_0)$, or to an alternative model $H_1 : y \sim F(y)$, according to the realizations of a latent binary indicator variable $\gamma \sim \text{Bern}(\omega)$, i.e.

$$y|\omega \sim \omega F(y; \theta_0) + (1 - \omega) F(y). \quad (1)$$

In Bayesian nonparametrics, it is common to assume a Dirichlet process mixture (DPM) model on $F(y)$, in order to describe the heterogeneity of realizations under the alternative model (Ferguson, 1983), i.e.

$$F(y) = \int F(y|\theta) G(d\theta), \quad (2)$$

where $G \sim DP(\alpha, G^*)$ indicates a Dirichlet Process, with concentration (mass) parameter $\alpha > 0$ and base distribution G^* . The null distribution $F(y; \theta_0)$ may be either parametric or nonparametric. For example, Do et al. (2005) and Guindani et al. (2014) employ DPM priors to infer differential gene expression (respectively, sequence abundances) under two different conditions. As a further example, Scarpa and Dunson (2009) consider a mixture of a null parametric model and an alternative nonparametric contamination to model functional data. A different approach is presented in Kim et al. (2009), who directly assume $y \sim F(y)$ as in (2), with the base measure G^* modeled as a “spiked” mixture of a point-mass and a diffuse measure. Canale et al. (2017) refer to (1) as an “outer” mixture, as opposed to the “inner” spike-and-slab mixture prior of Kim et al. (2009), and show how two types of mixtures yield structurally different priors. In particular, model (1)–(2) does not coincide with the typical DPM model, due to the constraint that $y \sim F(y; \theta_0)$ with probability ω .

In this paper, we assume the outer mixture formulation (1), and we further allow model selection varying across space and time, by modeling the slab component $F(y)$ in (1) with a flexible non-exchangeable species sampling prior process. More specifically, we assume a state-space model, where the state process is the realization of a conditionally identically distributed (CID) species sampling sequence (Berti et al., 2004; Bassetti et al., 2010). In this formulation, the state of a process at time t depends on past observations through a generalized Pólya urn scheme. This allows to describe the evolution of a non-stationary process over time, with limited assumptions on the temporal dependence among observations. With respect to Dirichlet processes, CID sequences only require that, conditionally upon the past, sequences of future observations are identically distributed, i.e. a weaker dependence than exchangeability and a requirement akin to predictive stationarity. For the analysis of regime-switching processes, CID sequences

provide a robust and flexible alternative to general hidden Markov and semi-Markov models (Airoldi et al., 2014).

We further account for spatial dependencies in the selection of the states, by coupling the species sampling model that guides the temporal evolution of the process with a spatial Markov random field (MRF). Thus, observations are assumed to have a higher probability *a priori* to originate from the baseline process if nearby locations also do. Similarly, deviations may propagate in space, although their levels could vary. Since the spatial MRF model is used at each time to infer if the process is in its baseline state, our approach is different than the one by (Jo et al., 2017), where a conditionally autoregressive (CAR) specification is used to specify the weights of a random probability measure, with no baseline and with independent observations at each location.

In an application to the detection of outbreaks of pneumonia and influenza (P&I) mortality in the continental United States, we show how our approach provides a more flexible description of spatio-temporal dependencies with respect to other methods used in disease surveillance, namely the thresholding-based method of Muscatello et al. (2008), which is similar to the one currently employed by the CDC, and Bayesian Markov-switching models, which assumes two levels of the process (Martínez-Beneito et al., 2008). We also compare the performance of our approach with “inner” spike-and-slab Dirichlet process (SS-DP) and Pitman-Yor (SS-PY) mixture priors, which enable clustering of observations without any assumption on their temporal dependence (Canale et al., 2017). Our approach allows to cluster observations at different states over time, avoiding a strict homogeneous Markov assumption, while still retaining spatial dependence in the modeling of outbreaks in neighboring regions.

The paper is organized as follows. In Section 2, we formulate our dynamic model selection approach, with a spiked CID process prior to capture the evolution of the process over time and a Markov Random Field prior to inform spatial selection. In Section 2.3, we describe posterior inference under a compound decision theoretic framework. In Section 3 we apply our model to a CDC pneumonia and influenza mortality dataset, and compare with respect to alternative methods. In Section 4, we further investigate the performance of our modeling approach by means of a simulation study. We conclude with some final remarks in Section 5.

2 Bayesian dynamic model selection

Let $Y(s, t)$ denote a spatio-temporal stochastic process defined on a spatial domain D , at times $t \in \mathcal{T}$. We have measurements available at a finite set of locations, $s \in \{s_1, \dots, s_n\} \subset D$, at discrete times $t = 1, \dots, T$. In the following, we assume D as a spatial lattice, although our framework may be extended to include spatial dependences on a general domain $D \subset \mathbb{R}^d$, $d > 1$. We assume a random effect formulation for the stochastic process $Y(s, t)$, i.e.

$$Y(s, t) = \mu(s, t) + \theta(s, t) + \epsilon(s, t), \quad (3)$$

where $\mu(s, t)$ denotes a general regression term, capturing the effect of baseline characteristics, which may include temporal trends and spatially varying covariates, $\theta(s, t)$ is

a pure random effect, capturing deviations from the baseline trend localized in space and time, and $\epsilon(s, t)$ is a residual uncorrelated measurement error, e.g.

$$\varepsilon(s, t) \stackrel{i.i.d.}{\sim} N(0, \tau_s^2), \quad (4)$$

for $s \in D$ and $t = 1, \dots, T$. The regression term $\mu(s, t)$ is modeled on the basis of available covariate information, according to the formulation which is typical in Bayesian spatio-temporal process modeling, see e.g. Banerjee et al. (2014). Here, we are mainly interested in inference about the random effect process, $\theta(s, t)$.

2.1 Spiked CID process prior

We formulate the dynamic model selection problem by assuming that at each location s and at each time point t , the random effect $\theta(s, t)$ is either a realization from a latent spatio-temporal baseline (“null”) process, e.g. $\theta(s, t) = \theta_0(s, t)$, or a realization from an alternate spatio-temporal process, i.e. $\theta(s, t) \sim G(s, t)$, with $G(s, t)$ indicating the marginal distribution of a general process $G(\cdot, \cdot)$ defined on $D \times \mathcal{T}$, which we further specify below. Without loss of generality, we assume $\theta_0(s, t) = 0$, i.e. under baseline conditions the observations are fully described by the regression $\mu(s, t)$, possibly including an intercept, and by the measurement error variances.

We further characterize the evolution of the process $\theta(\cdot, \cdot)$ over time as the realization of a sequence of conditional distributions, where the realization at time t , $\theta(s, t)$, depends on past realizations $\theta(s, j)$, $j < t$. More specifically, let $\boldsymbol{\theta}^{(t-1)}(s) = [\theta(s, 1), \theta(s, 2), \dots, \theta(s, t-1)]^T$ denote the $t-1$ -dimensional vector containing the past realizations of the latent process at location s , $s \in \{s_1, \dots, s_n\}$. Then, for $s \in \{s_1, \dots, s_n\}$ and any $t > 1$, we define the conditional distribution of $\theta(s, t)$ as the spiked mixture,

$$\theta(s, t) | \boldsymbol{\theta}^{(t-1)}(s) \sim (1 - \omega(s, t)) \delta_0(\cdot) + \omega(s, t) G_{t-1}(\cdot), \quad (5)$$

where $\delta_0(\cdot)$ denotes a point mass on the baseline value $\theta(s, t) = 0$, and $\omega(s, t)$ is a weight, which in general is assumed to be both spatially and temporally varying. The conditional distribution $G_{t-1}(\cdot)$ encodes the temporal dependence of the realizations of the process $\theta(s, t)$, given $\boldsymbol{\theta}^{(t-1)}(s)$. Here, we assume that the sequence of conditional distributions $G_{t-1}(\cdot)$, $t > 1$, is defined by the predictive distributions of a generalized species sampling sequence. For notational simplicity, we indicate with $G_{t-1}(\cdot)$ also the sequence of conditional cumulative distribution functions, and write:

$$G_{t-1}(x) = \Pr \left\{ \theta(s, t) \leq x | \boldsymbol{\theta}^{(t-1)}(s) \right\} = \sum_{j=1}^{t-1} p_{s, t-1}^{(j)} \delta_{\theta(s, j)}(-\infty, x] + r_{s, t-1} G^*(x), \quad (6)$$

where the weights $p_{s, t-1}^{(j)}$ specify the probability that the value of the process at time t coincides with a past value, $\theta(s, j)$, $j < t$, and $r_{s, t-1}$ is the remainder probability that $\theta(s, t)$ does not cluster with any previous value. $G^*(x)$ is a non-atomic base distribution, characterizing the values of the latent random effect under the alternate process. Thus, at each location we have $\sum_{j=1}^{t-1} p_{s, t-1}^{(j)} < 1$ and $r_{s, t-1} = 1 - \sum_{j=1}^{t-1} p_{s, t-1}^{(j)}$.

Due to the positive probability of ties $\theta(s, t) = \theta(s, j)$, $j < t$, equation (6) can be used to describe the temporal persistence of outbreaks (i.e., departures from the baseline process), and to identify sparse clusters of observations over time. To this purpose, we further assume that $G_{t-1}(\cdot)$ is a conditionally identically distributed (CID) species sampling sequence (Bassetti et al., 2010). The species sampling weights are defined as a function of independent (not necessarily identically distributed) latent random variables $W_1(s, t), W_2(s, t), \dots$ as follows:

$$\begin{aligned} p_{s,t-1}^{(j)} &= (1 - W(s, j)) \prod_{i=j+1}^{t-1} W(s, i), \\ r_{s,t-1} &= \prod_{i=1}^{t-1} W(s, i), \end{aligned} \tag{7}$$

with the realizations $W(s, j) \in (0, 1)$, $j = 1, \dots, t$. Thus, the prior predictive spatio-temporal effects $\theta(s, t+k)$, $k \geq 0$, are identically distributed conditionally to the past, $\theta^{(t-1)}(s)$. The specification of the weights $p_{s,t-1}^{(j)}$ and $r_{s,t-1}$ in (6)–(7) leads to different characterizations of CID sequences. A computationally convenient specification assumes

$$W(s, j) \stackrel{ind.}{\sim} \text{Beta}(j, 1), \tag{8}$$

for $j = 1, \dots, t$, which defines a *Beta-GOS* sequence at each location (Airoldi et al., 2014). Since the predictive weights depend on the latent sequence $W(\cdot, \cdot)$, the evolution of the process does not depend on the estimation of a single transition matrix across all time points. Therefore, this choice provides an alternative to more complicated non-homogeneous Hidden Markov Models for modeling a non-stationary temporal process. In particular, although our approach assumes two states of the process, the outbreaks can be clustered at multiple levels. The number of clusters is estimated directly from the data, using the Bayesian nonparametric specification. Moreover, by specifying the latent random variables as in (8), we assume that the process is characterized by long memory properties, with high and persistent autocorrelation, on average. This might be appropriate to capture the recurrent seasonality of the clusters over time. We refer to Airoldi et al. (2014) for details and further discussion of the asymptotic properties.

2.2 Modeling spatial selection

The model described in the previous section can be regarded as a *zero-inflated* CID sequence at each location s , which allows for unsupervised clustering of the null and alternative processes over time. We further allow for outbreaks to be clustered in space, by introducing a latent indicator at each $s \in \{s_1, \dots, s_n\}$ and $t = 1, \dots, T$,

$$\gamma(s, t) = \begin{cases} 0 & \text{if } \theta(s, t) = 0, \\ 1 & \text{if } \theta(s, t) \sim G^*(\cdot), \end{cases} \tag{9}$$

which specifies if the observations are a realization from the null or alternate process, respectively. Then, the probability of observing an outbreak at time t is a function of

the weights of the spiked mixture of CID sequence defined in (5)–(6). More specifically, we can compute the conditional probability of an outbreak at time t given the past realizations of the process, $\boldsymbol{\theta}^{(t-1)}(s)$, as

$$Pr(\gamma(s, t) = 1 | \boldsymbol{\theta}^{(t-1)}(s)) = \omega(s, t) \pi(s, t), \quad (10)$$

where $\pi(s, t) = \sum_{\{j: j < t, \gamma(s, j) = 1\}} p_{s, t-1}^{(j)} + r_{s, t-1}$ denotes the total probability mass associated to pairing $\theta(s, t)$ with any of the past realizations of the random effect, $j < t$, or to a new draw from G^* . Let $\boldsymbol{\Gamma}$ be the $n \times T$ matrix of the latent binary indicators $\gamma(s, t)$. Since (6) is a CID sequence, then, conditionally on $\boldsymbol{\Gamma}$, the marginal model of $\theta(s, t)$ is a spike-and-slab mixture,

$$\theta(s, t) | \gamma(s, t) \sim (1 - \gamma(s, t)) \delta_0(\cdot) + \gamma(s, t) G^*(\cdot).$$

In order to account for spatial dependence, in our application to outbreak detection, we leverage the lattice structure of the data, and model the weights $\omega(s, t)$ in (10) using a Markov random field prior (MRF). MRF priors have been widely used to model spatial dependence for non-dynamic variable selection (see, for example Zhang et al., 2014, for an application in neuroimaging). More specifically, we define a neighbor set of each location s , $N_{s,t}$, which in general could depend on the specific time t , and allow the probability of an outbreak to increase if some outbreaks are detected at the same time in neighboring locations. Let $\boldsymbol{\gamma}(N_{s,t}) = \{\gamma(k, t), k \in N_{s,t}\}$ denote the vector of selection indicators in $N_{s,t}$. Then, we can define the probability of an outbreak conditionally on the status of neighboring locations as

$$Pr(\gamma(s, t) = 1 | \boldsymbol{\theta}^{(t-1)}(s), \boldsymbol{\gamma}(N_{s,t})) = \omega^*(s, t) \pi(s, t),$$

where the weight $\omega^*(s, t)$ captures the spatial dependence in (10) conditionally on the neighborhood $N_{s,t}$ and it is defined as

$$\omega^*(s, t) = \frac{\exp(d + e \sum_{k \in N_{s,t}} \gamma(k, t))}{1 + \exp(d + e \sum_{k \in N_{s,t}} \gamma(k, t))}, \quad (11)$$

with $d \in \mathbb{R}$ being a sparsity parameter, and $e > 0$ indicating a smoothing parameter that controls the degree of spatial dependence in a neighborhood of s , $s \in \{s_1, \dots, s_n\}$. In particular, the conditional probability of an outbreak will be $(\exp(d)/1 + \exp(d))\pi(s, t)$, if $e = 0$, or if $\gamma(k, t) = 0$, for all $k \in N_{s,t}$, i.e. there are no outbreaks in a neighborhood of s at time t . Since $\omega^*(s, t)$ increases as a function of the number of locations $k \in N(s, t)$ where $\gamma(k, t) = 1$, our specification (11) favors the detection of spatial clusters of outbreaks. Other spatial priors could be considered, e.g. a spatial probit model where the probability of an outbreak depends on a latent gaussian Markov random field process. In Section 4, we provide a sensitivity analysis to guide the selection of the parameters d and e . In order to complete our model, we select an inverse-gamma conjugate prior for the region specific variance, i.e. $\tau_s^2 \sim IG(a_0, b_0)$, and we further choose a Normal distribution as the base measure for the atoms of the alternate process, that is $G^*(\cdot) \equiv \mathcal{N}(\mu_0, \sigma_0^2)$.

2.3 Posterior inference

In order to obtain inferences on the latent process $\theta(s, t)$ and the other parameters of the proposed dynamic selection process, we need to employ Markov chain Monte Carlo (MCMC) posterior sampling techniques. The details of the MCMC are provided in the supplementary material (Cassese et al., 2018).

The detection of outbreaks relies on the evaluation of the posterior probability of selection (PPS) of the alternate process at each location $s \in \{s_1, \dots, s_n\}$ and at each time $t = 1, \dots, T$, which is computed from the MCMC output for the elements of the selection matrix Γ as

$$PPS(s, t) \approx \frac{1}{B} \sum_{b=1}^B I_{\{\gamma^b(s, t) = 1\}}, \quad (12)$$

where $\gamma^b(s, t)$ denotes the value of the selection indicator $\gamma(s, t)$ sampled at iteration b out of B MCMC iterations after burn in.

Let $\mathbf{a} = \{a(s, t)\}$, $s \in \{s_1, \dots, s_n\}$, $t = 1, \dots, T$ be the vector of decisions (actions) in the multiple hypothesis testing problem, with $a(s, t) = 0$ indicating that the observations are assumed from the baseline process, and $a(s, t) = 1$ otherwise. In a compound decision theoretic framework, the spatio-temporal multiple testing problem for pointwise detection can be addressed by minimizing the posterior expectation of the loss function,

$$L(\Gamma, \mathbf{a}) = \sum_{s, t} \gamma(s, t) (1 - a(s, t)) + \lambda \sum_{s, t} (1 - \gamma(s, t)) a(s, t), \quad (13)$$

which weights false negatives and false positives results, through a penalty parameter, λ . The optimal decision rule minimizes $E(L(\Gamma, \mathbf{a})|\text{data})$ and corresponds to a threshold on the posterior probability of selection, i.e. $a^*(s, t) = 1$ if $PPS(s, t) > 1/(1 + \lambda)$, and $a^*(s, t) = 0$ otherwise (Sun et al., 2015). The value of λ can be chosen so to guarantee a small expected Bayesian FDR,

$$BFDR(h) = \frac{\sum_s \sum_t (1 - PPS(s, t)) \mathbf{I}_{\{(1 - PPS(s, t)) \leq h\}}}{\sum_s \sum_t \mathbf{I}_{\{(1 - PPS(s, t)) \leq h\}}}, \quad (14)$$

where $h = 1/(1 + \lambda)$ (Newton et al., 2004). For example, h may be chosen so that $BFDR(h) < 0.05$.

3 Detection of P&I outbreaks from CDC data

In this Section, we illustrate the performances of the proposed spiked CID process prior for the analysis of weekly data on pneumonia and influenza (P&I) mortality. The data are made publicly available by the U.S. National Center for Health Statistics (NCHS), a section of the Centers for Disease Control and Prevention (CDC) (<http://www.cdc.gov/nchs/>), and they are released weekly, in order to provide a system capable of supporting near real-time surveillance. Here, we focus on weekly percentage of P&I death in the continental U.S. states, over the years 2011–2014. This leads to a total of $n = 48$ regions

and $T = 208$ time points. P&I data are often used as a proxy measure for influenza activity and have been extensively employed for the detection of influenza epidemics. They are considered a reliable and specific endpoint for studying timing and amplitude of influenza related mortality, both at the local and national scales.

We also comment on the inference obtained by the thresholding approach in Muscatello et al. (2008), which is similar to the method currently employed by the CDC for their analyses. The method relies on a threshold approach, which is based on an estimate of a non-epidemic seasonal baseline. However, since the threshold is based on a nationally defined baseline, local patterns of influenza outbreaks may be difficult to detect. The model also assumes spatial and temporal independence of the outbreaks (Amorós et al., 2015), which may result in decreased power of the testing procedures. To overcome those shortcomings, Martínez-Beneito et al. (2008) have proposed a method based on hidden Markov switching models to identify and cluster epidemic periods over time, although with no spatial dependence. They also assume only two levels of the process, whereas our method allows to cluster outbreaks at multiple levels. We also consider two purely Bayesian nonparametric methods, namely the spike-and-slab Dirichlet Process prior (SS-DP) of Kim et al. (2009) and the spike-and-slab Pitmann-Yor prior (SS-PY) of Canale et al. (2017), which assume exchangeability of the data. A more comprehensive comparison on simulated data is presented in Section 4.

To provide a limited seasonal adjustment, we first center the P&I weekly percentage of death data with respect to their yearly mean. We then fit the spiked CID model (3)–(7) by setting $G^*(\cdot) \equiv N(0, 25)$, $\tau_s \sim IG(a_0 = 2.001, b_0 = 0.04)$, corresponding to a vague specification centered around a mean of 0.04, and the Beta distribution for the weights as in (7). Lastly, we set the parameters of the MRF in (11) to $d = -1$ and $e = 0.5$, based on the sensitivity analysis presented in Section 4. In the implementation of the SS-DP and SS-PY models, we set the base measure as a weighted mixture of a point mass at zero and a $N(0, 25)$, with a $Beta(1, 1)$ prior on the mixture weight. Additionally, we set the concentration parameter to 1, and set the discount parameter of the Pitman-Yor process to 0.5. The results reported below are obtained by running MCMC chains with 10,000 iterations and a burn-in of 2,000. Using a 4-core 2.5GHz [®]Intel core i5 processor with 8 GB memory, our code takes approximately 90 minutes to run. We assessed convergence by inspecting the MCMC traces, and more formally by using the Geweke (1992) and Heidelberger and Welch (1981) tests. As an example, the z-score from the Geweke (1992) test was -0.01957 for Γ , whereas the Heidelberger and Welch (1981) test returned a p-value of 0.835.

We first illustrate our inference on the estimation of the matrix Γ , obtained as outlined at the end of Section 2.3, and the values of the process $\theta(s, t)$. Figure 1(a) shows a choropleth map of the P&I mortality rates for the time window between the 47th week of 2012 and the 3rd week of 2013. The map provides an illustration of the spatio-temporal evolution of P&I mortality rates. For example, the last row shows higher intensities of P&I mortality across all states, with nearby States characterized, on average, by similar values. Figure 1(b) shows the results of our inference, namely the outbreaks detected by our method, after thresholding the PPS using a BFDR control at the $\alpha = 0.05$ level. The states identified as under an epidemic state are colored in black, whereas the

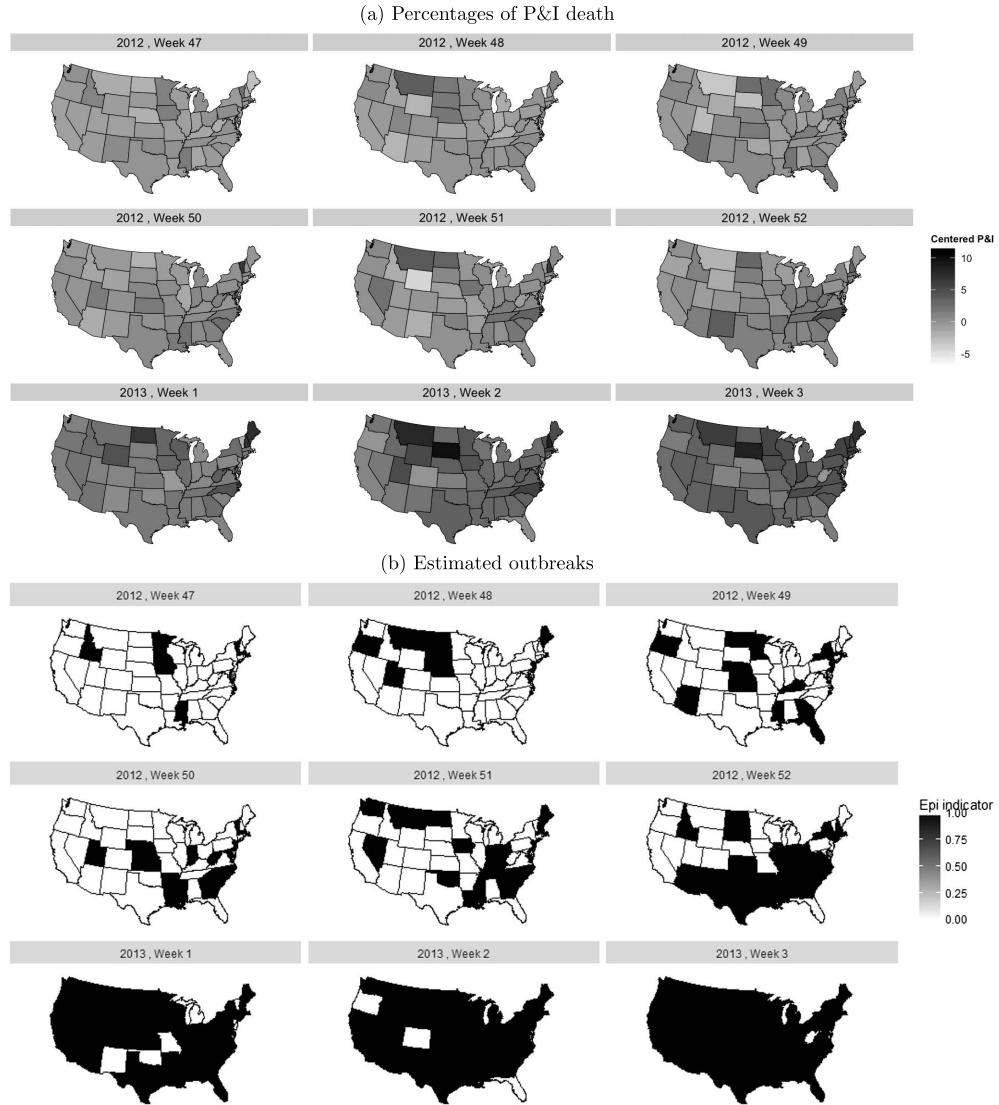


Figure 1: Case Study: (a) Map of centered P&I mortality rates over the time window: week 47 in 2012 – week 3 in 2013; (b) outbreaks detected by our method, using a 0.05 Bayesian FDR threshold of the PPS. States marked in black are identified as epidemic.

baseline process is marked in white. The figure highlights the spatio-temporal patterns and spread of diseases typical of an epidemic period: epidemics start in a few isolated States at first, possibly with limited temporal persistence (weeks 47–50), then there's a more noticeable propagation, in a behavior typical of transitioning periods (weeks 51–52), until all or almost all the States have been involved (weeks 1–3).

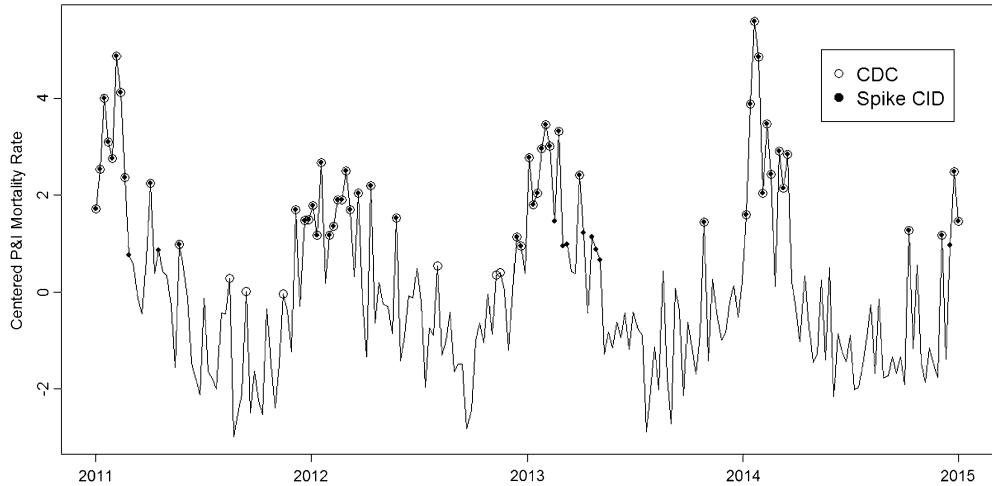


Figure 2: Case Study: Weekly P&I mortality rates for the state of Oklahoma, data centered accordingly to their yearly mean (2011–2014). Black dots (\bullet) indicate outbreaks identified by the proposed spike CID process, empty circles (\circ) by the CDC thresholding method.

As a further illustration of the temporal dependence, Figure 2 shows the centered time series of the weekly P&I mortality rates for the state of Oklahoma over 2011–2014. The black dots indicate the outbreaks flagged by our model, whereas the results of the CDC thresholding approach are indicated by empty circles. For the sake of graphical clarity, we report the plots for the SS-DP and SS-PY in the supplementary material, and only briefly discuss the comparison here. All methods identify outbreaks in the Winter, and more specifically close to the end of one year and the beginning of the next. Indeed, it is known that P&I epidemics usually start in this period, with patterns that are quite irregular and with unpredictable persistence over time, as highlighted also by our results. The method employed by CDC identifies also some isolated time points, which are not detected by any of the other methods. Those isolated peaks occur at times generally characterized by low P&I mortality, which may be due to the use of a national baseline for threshold determination, possibly affecting the sensitivity of the method. On the other hand, SS-DP and SS-PY share similar performances and tend to select only higher values of P&I mortality rates. Based on the simulation study in Section 4, we speculate this might be due to the fact that those methods do not account for the spatio-temporal dependence between observations. Instead, our method clearly identifies temporally persistent clusters of P&I outbreaks, which may provide relevant information to delineate the duration of the outbreaks.

We further compared our results with those obtained fitting a Markov switching model on each series of differenced incidence rates, following the guidelines in Martínez-Beneito et al. (2008). More specifically, we set 0.001 and 25 as respectively the lower and upper bounds to specify the uniform distribution for the variance parameters in

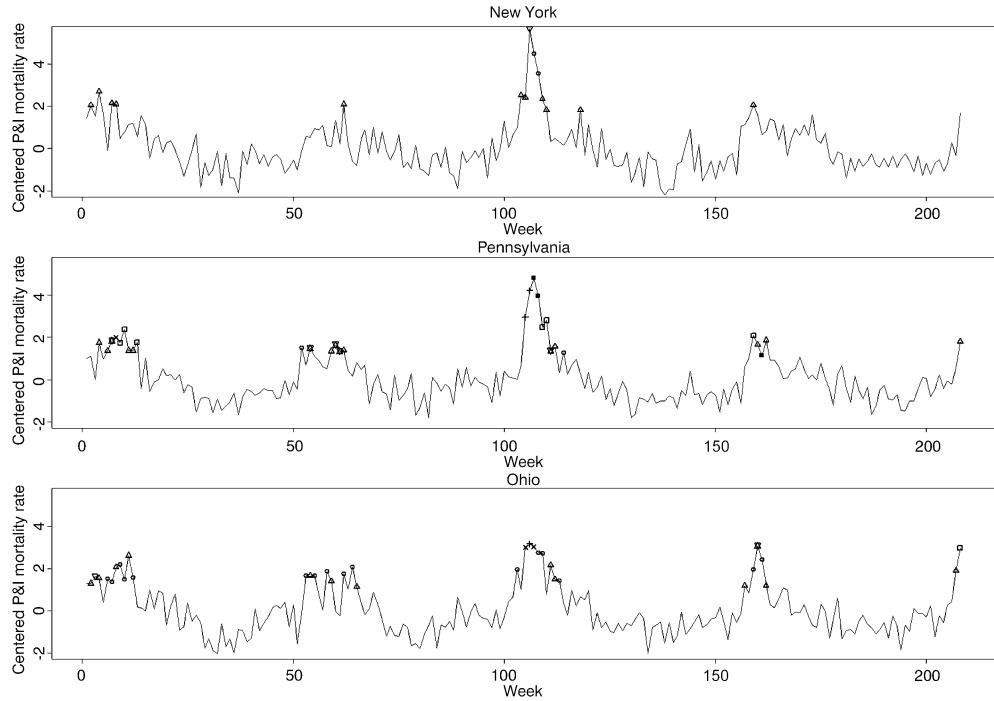


Figure 3: Case study: Time series for New York, Pennsylvania and Ohio. Aberrant mortality rate cluster labels are marked with different shapes within each time series.

the model. The result of this analysis led to flag all or none of the weeks as epidemic, depending on the State. This unrealistic result is probably related to Martínez-Beneito et al. (2008) modeling the differenced time series as a mixture of two hidden dynamics. The first dynamic is of lower variance and corresponds to low incidence rate levels, while the other one is of relatively higher variance and corresponds to high incidence rate levels. Although this assumption may be appropriate to describe the rates of influenza-like illness data, it seems not to be realistic to capture the behavior of the P&I mortality data.

Lastly, Figure 3 shows the effect of modeling spatial dependence on within-series clustering, for the States of New York, Pennsylvania and Ohio. Each cluster partition is obtained by post-hoc analysis of the MCMC iterations. For the sake of illustration, here we show the partitions obtained from the last MCMC iteration. Alternatively, one could find a modal partition (Kim et al., 2009). Each cluster is represented by a different shape in Figure 3. Observations assigned to the same cluster show similar mortality rate levels, with clear temporal persistence. Also, the clusters of those time points characterized by high mortality rates (i.e., significantly above the yearly mean) have similar patterns across the three neighboring regions.

4 Simulation study

In this Section, we assess the performance of our model on a set of simulated scenarios, and compare the results with the threshold based model of Muscatello et al. (2008), which is similar to the method currently employed by the CDC, the Markov switching model of Martínez-Beneito et al. (2008), the SS-DP of Kim et al. (2009) and the SS-PY of Canale et al. (2017). We assume \mathcal{D} as the spatial lattice comprising the 48 US continental States and consider two data generating mechanisms, mimicking the type of spatio-temporal dependence of epidemic incidence data.

Data Generating Processes: First, we consider a deterministic model of spatio-temporal propagation of epidemic outbreaks across the US, and the corresponding recovery of a non-epidemic situation (baseline). More specifically, we assume that a process may spring from the North West of the United States, say the state of Washington, at time t , based on a first order HMM with only two latent states, characterizing the epidemic and non-epidemic processes. At time $t+1$, the process originated in the state of Washington spreads to nearby states, say Oregon and Idaho, following the pattern defined by state borders. The process then propagates similarly in neighboring states at times $t+2, t+3, \dots$. This deterministic process generates a strong spatio-temporal dependence in the selection of the baseline and alternate processes.

The second scenario is designed to simulate more localized spatial dependence. Specifically, we subdivide all the continental US states into nine divisions (New England, Middle Atlantic, South Atlantic, East South Central, West South Central, East North Central, West North Central, Mountain, and Pacific), using the R function `state.division`. For each of those divisions, we simulate a time series of binary indicators $\gamma(\cdot, t)$, according to a two states HMM. We assume the same realizations $\gamma(s, t)$, $t = 1, \dots, T$, for all states s in a division. Yet, we allow for a range of different distributions to generate each observation $y(s, t)$ under either the baseline model f_0 ($\gamma(s, t) = 0$) and the alternate model f_1 ($\gamma(s, t) = 1$), as we detail further below. We set the transition probability matrix of the HMM modeling the change points between epidemic and non-epidemic processes (only in the state of Washington in the first scenario; in each division in the second scenario) to

$$\begin{aligned} \gamma(s, t) = 0 & \quad \gamma(s, t) = 1 \\ \gamma(s, t-1) = 0 & \quad \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}, \\ \gamma(s, t-1) = 1 & \end{aligned}$$

with initial condition $\gamma(s, 1) = 1$. The diagonal elements contain the probabilities of persistence, which are assumed relatively higher for the epidemic than for the non-epidemic process. In particular, we assume a low probability of transitioning to an epidemic state, $Pr(\gamma(s, t) = 1 | \gamma(s, t-1) = 0) = 0.1$, reflecting the irregular nature of epidemic periods.

In both scenarios, the observations are generated independently for each state according to the following procedure. If $\gamma(s, t) = 0$, we assume $y(s, t) | (\gamma(s, t) = 0) \stackrel{i.i.d.}{\sim} f_0 \equiv N(0, \tau_{sim}^2)$. On the other hand, if $\gamma(s, t) = 1$, we assume $y(s, t) | (\gamma(s, t) = 1) \stackrel{i.i.d.}{\sim}$

$f_1 \equiv \sum_{l=1}^L c_l N(\mu_l, \tau_{sim}^2)$, i.e. a mixture of L components with weights c_l , $l = 1, \dots, L$. More precisely, here we show the results for a four component mixture of normal distributions, $f_1 \equiv 0.25N(-1, \tau_{sim}^2) + 0.25N(-0.5, \tau_{sim}^2) + 0.25N(0.5, \tau_{sim}^2) + 0.25N(1, \tau_{sim}^2)$. We further set $\tau_{sim}^2 = 0.04$ and the number of time points $T = 200$. We note that in both scenarios the data generating processes are different than the proposed model, and allow for both higher and lower process means with respect to the baseline. Furthermore, the mixture framework allows to cluster the realizations of the process, without imposing any temporal persistence to such clusters, contrary to the assumptions of our model (6).

Model Fitting: We employ our Spiked Process prior model (6)–(8). We use a vague specification for the base measure, namely $G^*(\cdot) \equiv N(0, 5^2)$. We choose a vague inverse gamma distribution for the variance of the measurement error (3), i.e. $\tau_s^2 \sim \text{IG}(a_0 = 2.001, b_0 = 0.04)$, which leads to a prior mean of 0.04 with relatively large variance. Unless otherwise noted, we present here the results obtained by informing spatial dependence via the MRF prior with $d = -1$ and $e = 0.5$. We ran 10,000 iterations of the MCMC scheme, discarding the first 1,000 iterations as burn-in. We assessed convergence by inspecting the MCMC sample traces for all parameters. We further tested the convergence and stationarity of the chains by the Geweke (1992) diagnostic test and the Heidelberger and Welch (1981) approach, respectively.

Posterior inference and sensitivity analysis: Figure 4(a) shows the map of the posterior probabilities of selection $PPS(s, t)$ for one of the simulated datasets in the first scenario, across the 48 US continental states and for 6 representative time points (100–105). The maps illustrate the spread of an epidemic process both spatially and temporally. Figure 4(b) shows the corresponding map of binary indicators $\gamma(s, t)$ of the true spatio-temporal processes over the same period, with black-colored states indicating $\gamma(s, t) = 1$. Our method appears to recover the true spatio-temporal pattern of the processes quite closely. Some results for the second scenario are reported in the Supplementary material.

Next we investigate the effect of the choice of the MRF parameters d and e on the selection performances. It is well known that allowing e to vary widely can lead to a phase transition problem, i.e. the expected number of selections increases greatly for small increments of e . This is warranted by (11) being an increasing function of the number of indicators $\gamma(s, t)$ equal to 1. Therefore, a careful selection of the parameter e is necessary to avoid inaccurate inferences, due to too many false selections. Li et al. (2015) show that e should be small when the average number of neighbors is large; viceversa, a small d is required to induce small prior odds of $\theta(s, t) \sim f_1$ for a large number of sites. Indeed, for $e = 0$, the MRF prior reduces to an independent Bernoulli, with parameter $\exp(d)/[1 + \exp(d)]$, and thus can be used to provide a lower bound on the prior probability of selection. As for e , any value of e below the phase transition point can be considered a good choice, with values closer to the phase transition point leading to higher prior probabilities of selection in a neighborhood of sites already assigned to the alternate process.

Here, we report the effect of different choices of the parameters d and e in terms of the resulting specificity, sensitivity and accuracy rates. For each s and t , sensitivity

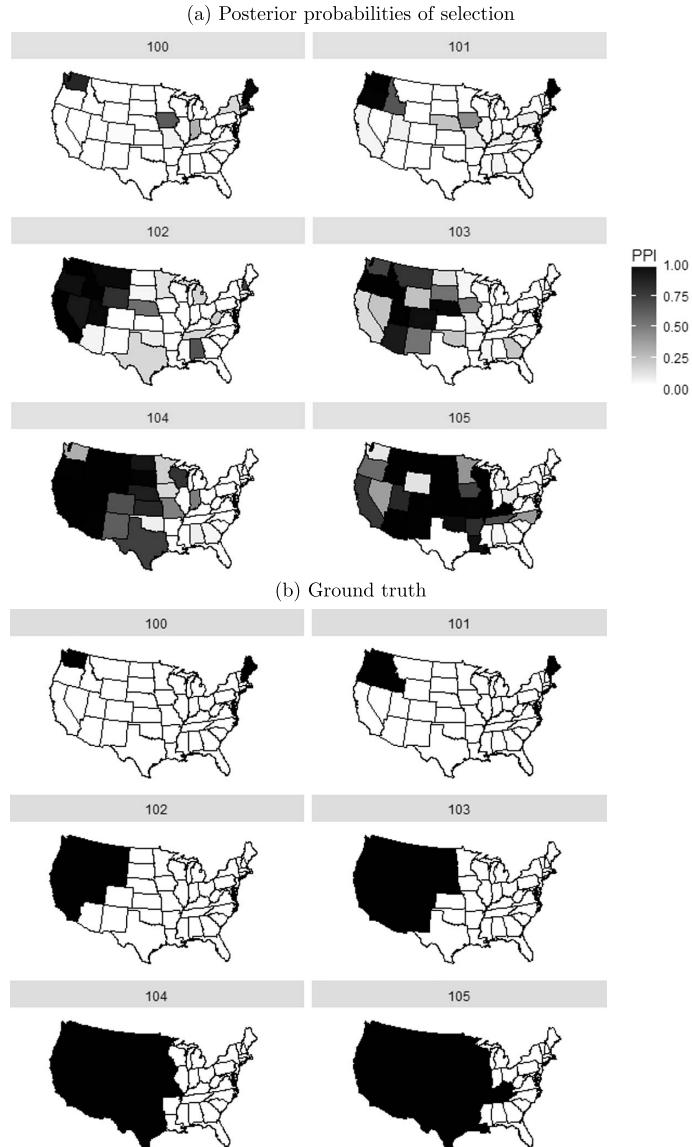


Figure 4: Simulation Study: (a) Map of posterior probabilities of selection $PPS(s,t)$ in a simulated dataset from the first scenario in the simulation study in Section 4, over the time window (time points 100–105); (b) map of true values of $\gamma(s,t)$ in the first scenario over the time window from point 100 to 105.

is calculated as the ratio of true positive selections (TP) over the total number of observations truly from the alternate process, sensitivity as the ratio of true negative selections (TN) over the total number of observations truly from the baseline process,

		$e = 0$	$e = 0.4$	$e = 0.5$	$e = 0.6$	$e = 0.7$
Scenario 1	Sensitivity	0.6914 (0.0543)	0.7381 (0.0511)	0.7421 (0.0492)	0.7440 (0.0464)	0.7435 (0.0445)
	Specificity	0.9922 (0.0033)	0.9898 (0.0044)	0.9895 (0.0045)	0.9894 (0.0044)	0.9895 (0.0041)
	Accuracy	0.8922 (0.0175)	0.9068 (0.0130)	0.9078 (0.0132)	0.9082 (0.0140)	0.9078 (0.0147)
Scenario 2	Sensitivity	0.7085 (0.0257)	0.7485 (0.0230)	0.7516 (0.0223)	0.7510 (0.0224)	0.7508 (0.0227)
	Specificity	0.9929 (0.0018)	0.9905 (0.0025)	0.9902 (0.0026)	0.9901 (0.0025)	0.9901 (0.0024)
	Accuracy	0.8995 (0.0098)	0.9111 (0.0084)	0.9118 (0.0087)	0.9115 (0.0097)	0.9114 (0.0096)

Table 1: Simulation Study: sensitivity analysis to the choice of e for the two scenarios averaging over 20 independently generated datasets with $d = -1$, standard error within brackets.

and accuracy is the ratio of observations correctly assigned to either the baseline or alternate processes (TP+TN) over the total number of observations. Table 1 shows the average and standard error of the previous measures, obtained for each of the two scenarios over 20 independently generated datasets, for varying values of e with $d = -1$ fixed. As expected, the table suggests that greater values of e generally lead to increased sensitivity at the expense of the specificity. In the first scenario, setting $e = 0.6$ leads to the best performance in terms of accuracy with $e = 0.5$ having similar performance, whereas in the second scenario the best performance is obtained for $e = 0.5$. Those results suggest that an optimal setting $0.5 \leq e \leq 0.6$, and we discuss the case $e = 0.5$ for the results presented in this manuscript. For $e = 0$, the model assumes no spatial dependence, and deviations from the baseline process are detected only based on each single time series, since the MRF reduces to a Bernoulli prior. We observe markedly increased sensitivity and accuracy for considering the spatial information in the selection prior, with only a small reduction in specificity. We also explored the effect of different choices of the sparsity parameters d , for which accuracy appears to be maximized at $d = -1$. (see Table 2).

Performance comparisons: We further compare the performances of our method with the thresholding method used by CDC (Muscatello et al., 2008), the Markov Switching model in Martínez-Beneito et al. (2008) (using the public open source code of Conesa et al., 2009) and the Bayesian nonparametric methods, SS-DP and SS-PY. The comparison results are averaged over the 20 independently generated datasets. The thresholding procedure of Muscatello et al. (2008) achieves high specificity, but performs much worse in terms of sensitivity, and accuracy: in the first scenario, the CDC procedure achieves 0.4245(0.0678) sensitivity, 0.9998(0.0002) specificity, and 0.7993(0.0635) accuracy; in the second scenario, 0.4273(0.0369) sensitivity, 0.9997(0.0004) specificity, and 0.8102(0.0275) accuracy. The Markov switching model performs quite well, comparably

		$d = 0$	$d = -1$	$d = -2$
Scenario 1	Sensitivity	0.5548 (0.1194)	0.7421 (0.0492)	0.7274 (0.0602)
	Specificity	0.9967 (0.0015)	0.9895 (0.0045)	0.9932 (0.0035)
	Accuracy	0.8406 (0.0725)	0.9078 (0.0132)	0.9059 (0.0120)
Scenario 2	Sensitivity	0.5598 (0.0709)	0.7516 (0.0223)	0.7319 (0.0267)
	Specificity	0.9966 (0.0010)	0.9902 (0.0026)	0.9937 (0.0019)
	Accuracy	0.85134 (0.0331)	0.9118 (0.0087)	0.9078 (0.0087)

Table 2: Simulation Study: sensitivity analysis to the choice of d for the two scenarios averaging over 20 independently generated datasets with $e = 0.5$. Standard errors are reported in parentheses.

with our model. This is in line with our expectations, since the data are simulated from a HMM with two states. The performance of the Markov switching model slightly decreases in the presence of stronger spatial dependence: in the first scenario, the model achieves 0.7747(0.0931) sensitivity, 0.9832(0.0018) specificity, and 0.9046(0.0423) accuracy; in the second scenario, the achieved performance is 0.7905(0.0869) sensitivity, 0.9820(0.0041) specificity and 0.9164(0.0246) accuracy. However, the Markov switching model assumes the differenced time series are realizations from a mixture with two hidden dynamics, characterized by low and high variance. This may limit the ability of the model to detect outbreaks, if more than two states are reasonably supported by the data, as in our data analysis. The results from the comparison with the two exchangeable Bayesian nonparametric models confirms the conclusions from the data analysis: failing to account for spatio-temporal dependence leads to a loss in sensitivity, despite maintaining high specificity. Specifically, in the first scenario, the SS-DP achieves 0.6526(0.06163) sensitivity, 0.9935(0.0019) specificity and 0.8768 (0.0359) accuracy, whereas the SS-PY achieves 0.66051(0.0625) sensitivity, 0.9932(0.001) specificity, and 0.8810(0.0222) accuracy. In the second scenario, the SS-DP achieves 0.6725(0.0451) sensitivity, 0.9930(0.0021) specificity and 0.8869(0.0209) accuracy, whereas the SS-PY achieves 0.6835(0.0322) sensitivity, 0.9922(0.0010) specificity and 0.8879(0.012) accuracy. Those performances are slightly worse than those achieved by our Spiked-CID process with no-spatial dependence ($e = 0$, see Table 1). Therefore, we are led to conclude that borrowing information across time and sites leads to more stable inference and higher accuracy for the Spike-CID process with respect to exchangeable Bayesian non-parametric priors.

Prediction: Outbreak detection typically relies on the nearly real-time monitoring of carefully chosen proxy measures, and not on the prediction of future values of the observed process. This is due to the difficulty of identifying the early determi-

nants/predictors of an outbreak in syndromic surveillance. For example, P&I data are usually considered as proxy measures for influenza activity. The analysis of past data is important for monitoring disease trends. At the same time, detection of past outbreaks is meaningful in evaluating the effectiveness of disease control programs and policies. Thus, most models typically employed for outbreak detection do not focus on statistical prediction. Nevertheless, here we investigate the prediction performance of the Spiked-CID process in a leave-one-out simulation, where we assume the value of the process at the last available time-point, $y(s, T)$, is not observed. More in detail, let $\gamma^* = \{\gamma^*(s)\}$ indicate the vector comprising the true epidemic status at each location at time T . Then, we can compute the posterior predictive probability of assigning the correct epidemic status,

$$p(\gamma(s, T) = \gamma^*(s) | y(s, 1), \dots, y(s, T-1)) = \int p(\gamma(s, T) = 1 | \boldsymbol{\theta}^{(t-1)}(s)) \\ \times p(\boldsymbol{\theta}^{(t-1)}(s) | y(s, 1), \dots, y(s, T-1)),$$

using Monte Carlo composition sampling from the MCMC posterior samples. Results are obtained for the 48 states, averaged across 20 simulated datasets, according to the data generating mechanisms previously outlined. The assessment about the epidemic status is conducted by thresholding the posterior predictive probability at the median, i.e. we set $\hat{\gamma}(s, T) = 1$ if $p(\gamma(s, T) = 1 | y(s, 1), \dots, y(s, T-1)) > 0.5$. In the first simulation scenario, our method correctly detects the true status 69.8% of the times (median across all datasets, IQR 40.6% – 98.4%); in the second simulated scenario, which describes more localized patterns of spatial dependence, the method correctly detects the true status 62.5% of the times (median across all datasets, IQR 55.2% – 83.3%).

5 Discussion

We have proposed a Bayesian nonparametric approach for modeling temporary and localized changes in the distribution of a spatio-temporal process. Our proposal defines a spiked process prior, where clustering-inducing CID species sampling prior processes and a Markov random field prior are used to identify and propagate changes over time and space. We show good performances of our methods in the analysis of CDC data on P&I mortality rates and in a simulation study. Several generalizations of the proposed modeling framework are possible. The Markov random field prior formulation is particularly suitable for the analysis of spatio-temporal areal data. For point-referenced data, spatial selection could be obtained via a latent logistic regression with a Gaussian spatial random effect. Alternatively, it could be possible to incorporate spatial dependence in a more direct way, by defining interacting partially CID sequences (Fortini et al., 2017). Scalability of the MCMC posterior sampling is a limitation of the proposed method. However, CID sequences provide a promising flexible framework for scalable and fast implementation through sequential Monte Carlo approaches and approximate MCMC algorithms. Those extensions are the objective of ongoing work.

Supplementary Material

Supplementary Material to “A Bayesian Nonparametric Spiked Process Prior for Dynamic Model Selection” (DOI: [10.1214/18-BA1116SUPP](https://doi.org/10.1214/18-BA1116SUPP); .pdf).

References

Airoldi, E., Costa, T., Bassetti, F., Guindani, M., and Leisen, F. (2014). “Generalized Species Sampling Priors with Latent Beta reinforcements.” *Journal of the American Statistical Association*, 109(508): 1466–1480. [MR3293604](https://doi.org/10.1080/01621459.2014.950735). doi: <https://doi.org/10.1080/01621459.2014.950735>. 555, 557

Amorós, R., Conesa, D., Martínez-Beneito, M., and López-Quílez, A. (2015). “Statistical methods for detecting the onset of influenza outbreaks: a review.” *Revstat – Statistical Journal*, 13(1): 41–62. [MR3342204](https://doi.org/10.1344/revstat2015.0). 560

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. Boca Raton, London: Chapman & Hall/CRC, 2nd edition. [MR3362184](https://doi.org/10.1201/b13362). 556

Bassetti, F., Crimaldi, I., and Leisen, F. (2010). “Conditionally identically distributed species sampling sequences.” *Advances in Applied Probability*, 42(2): 433–459. [MR2675111](https://doi.org/10.1239/aap/1275055237). doi: <https://doi.org/10.1239/aap/1275055237>. 554, 557

Berti, P., Pratelli, L., and Rigo, P. (2004). “Limit Theorems for a Class of Identically Distributed Random Variables.” *The Annals of Probability*, 32(3): 2029–2052. [MR2073184](https://doi.org/10.1214/009117904000000676). doi: <https://doi.org/10.1214/009117904000000676>. 554

Canale, A., Lijoi, A., Nipoti, B., and Prünster, I. (2017). “On the Pitman–Yor process with spike and slab base measure.” *Biometrika*, 104(3): 681–697. [MR3694590](https://doi.org/10.1093/biomet/asw041). doi: <https://doi.org/10.1093/biomet/asw041>. 554, 555, 560, 564

Cassese, A., Zhu, W., Guindani, M., and Vannucci, M. (2018). “Supplementary Material to “A Bayesian Nonparametric Spiked Process Prior for Dynamic Model Selection”.” *Bayesian Analysis*. doi: <https://doi.org/10.1214/18-BA1116SUPP>. 559

Conesa, D., López-Quílez, A., Martínez-Beneito, M. A., Miralles, M. T., and Verdejo, F. (2009). “FluDetWeb: an interactive web-based system for the early detection of the onset of influenza epidemics.” *BMC Medical Informatics and Decision Making*, 9(1): 36. 567

Do, K., Müller, P., and Tang, F. (2005). “A Bayesian mixture model for differential gene expression.” *Journal of the Royal Statistical Society, Series C*, 54(3): 627–644. [MR2137258](https://doi.org/10.1111/j.1467-9876.2005.05593.x). doi: <https://doi.org/10.1111/j.1467-9876.2005.05593.x>. 554

Efron, B. (2004). “Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis.” *Journal of the American Statistical Association*, 99(465): 96–104. [MR2054289](https://doi.org/10.1198/016214504000000089). doi: <https://doi.org/10.1198/016214504000000089>. 554

Efron, B. (2008). "Microarrays, Empirical Bayes and the Two-Groups Model." *Statistical Science*, 23(1): 1–22. MR2431866. doi: <https://doi.org/10.1214/07-STS236>. 554

Ferguson, T. S. (1983). "Bayesian density estimation by mixtures of normal distributions." *Recent Advances in Statistics*, 24(1983): 287–302. MR0736538. 554

Fortini, S., Petrone, S., and Sporysheva, P. (2017). "On a notion of partially conditionally identically distributed sequences." *Stochastic Processes and their Applications*, 128(3): 819–846. MR3758339. doi: <https://doi.org/10.1016/j.spa.2017.06.008>. 569

Geweke, J. (1992). "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments." In *Bayesian Statistics*, 169–193. University Press. MR1380276. 560, 565

Guindani, M., Sepúlveda, N., Paulino, C. D., and Müller, P. (2014). "A Bayesian semi-parametric approach for the differential analysis of sequence counts data." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 63(3): 385–404. MR3238158. doi: <https://doi.org/10.1111/rssc.12041>. 554

Heaton, M., Banks, D., Zou, J., Karr, A., Datta, G., Lynch, J., and Vera, F. (2012). "A spatio-temporal absorbing state model for disease and syndromic surveillance." *Statistics in Medicine*, 31(19): 2123–2136. MR2956066. doi: <https://doi.org/10.1002/sim.5350>. 554

Heidelberger, P. and Welch, P. D. (1981). "A spectral method for confidence interval generation and run length control in simulations." *Communications of the ACM*, 24(4): 233–245. MR0611745. doi: <https://doi.org/10.1145/358598.358630>. 560, 565

Jo, S., Lee, J., Müller, P., Quintana, F. A., and Trippa, L. (2017). "Dependent Species Sampling Models for Spatial Density Estimation." *Bayesian Analysis*, 12(2): 379–406. MR3620738. doi: <https://doi.org/10.1214/16-BA1006>. 555

Kim, S., Dahl, D. B., and Vannucci, M. (2009). "Spiked Dirichlet process prior for Bayesian multiple hypothesis testing in random effects models." *Bayesian Analysis*, 4(4): 707–732. MR2570085. doi: <https://doi.org/10.1214/09-BA426>. 554, 560, 563, 564

Li, F., Zhang, T., Wang, Q., Gonzalez, M., Maresh, E., and Coan, J. (2015). "Spatial Bayesian variable selection and grouping for high-dimensional scalar-on-image regression." *The Annals of Applied Statistics*, 9(12): 687–713. MR3371331. doi: <https://doi.org/10.1214/15-AOAS818>. 565

Madigan, D. (2005). "Bayesian data mining for health surveillance." In *Spatial and Syndromic Surveillance for Public Health*, 203–221. John Wiley & Sons, Ltd. 553

Martínez-Beneito, M., Conesa, D., López-Quílez, A., and López-Maside, A. (2008). "Bayesian Markov switching models for the early detection of influenza epidemics." *Statistics in Medicine*, 27(22): 4455–4468. MR2528524. doi: <https://doi.org/10.1002/sim.3320>. 553, 555, 560, 562, 563, 564, 567

Muscatello, D., Morton, P., Evans, I., and Gilmour, R. (2008). “Prospective surveillance of excess mortality due to influenza in New South Wales: feasibility and statistical approach.” *Communicable Diseases Intelligence*, 32(4): 435–442. 553, 555, 560, 564, 567

Newton, M. A., Noueiry, A., Sarkar, D., and Ahlquist, P. (2004). “Detecting differential gene expression with a semiparametric hierarchical mixture method.” *Biostatistics*, 5(2): 155–176. 559

Scarpa, B. and Dunson, D. B. (2009). “Bayesian Hierarchical Functional Data Analysis Via Contaminated Informative Priors.” *Biometrics*, 65(3): 772–780. MR2649850. doi: <https://doi.org/10.1111/j.1541-0420.2008.01163.x>. 554

Sun, W., Reich, B. J., Tony Cai, T., Guindani, M., and Schwartzman, A. (2015). “False discovery control in large-scale spatial multiple testing.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1): 59–83. MR3299399. doi: <https://doi.org/10.1111/rssb.12064>. 559

Zhang, L., Guindani, M., Versace, F., and Vannucci, M. (2014). “A spatio-temporal non-parametric Bayesian variable selection model of fMRI data for clustering correlated time courses.” *NeuroImage*, 95: 162–175. 558

Zou, J., Karr, A., Banks, D., Heaton, M., Datta, G., Lynch, J., and Vera, F. (2012). “Bayesian methodology for the analysis of spatial-temporal surveillance data.” *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 5(3): 194–204. MR2929962. doi: <https://doi.org/10.1002/sam.10142>. 554