Bayesian Nonparametric Spatially Smoothed Density Estimation

Timothy Hanson, Haiming Zhou, and Vanda Inácio de Carvalho

March 19, 2018

Abstract

A Bayesian nonparametric density estimator that changes smoothly in space is developed. The estimator is built using the predictive rule from a marginalized Polya tree, modified so that observations are spatially weighted by their distance from the location of interest. A simple refinement is proposed to accommodate arbitrarily censored data and a test for whether the density is spatially varying is also developed. The method is illustrated on two real datasets, and an R function `SpatDensReg` is provided for general use.

1 Introduction

Geographic Information Systems (GIS) technology has exploded over the last several decades due to impactful advances in data storage, computing power, sophisticated processing techniques, and visualization software. Accordingly, there has been an increasing need for the development of state of the art statistical models for spatial data as well (for an overview of developed methods, see Gelfand, Diggle, Fuentes, and Guttorp, 2010). Much recent literature has focused on spatially-varying trends in the form of random fields. Although fundamental, spatially-varying density estimation has received much less attention, perhaps due to challenges inherent in adapting existing methods to the spatial setting. This paper is then motivated by the need to fill a particular gap in the literature: provide a conceptually simple and computationally feasible, yet competitive, approach to modeling spatially dependent distributions.

The field of Bayesian nonparametrics has offered several viable spatially varying density estimators over the last decade, the majority of which are based on convolutions of a continuous kernel with a spatially varying discrete measure. More specifically, all proposed methods have been extensions of Dirichlet process (DP) mixture models (Escobar and West, 1995) towards dependent Dirichlet process (DDP) mixtures (MacEachern, 2001). The first such contribution was proposed by Gelfand, Kottas,

and MacEachern (2005). These authors developed a DDP for point-referenced spatial data where the underlying DP base measure was taken to be a Gaussian process over Euclidean space; as a result, the density estimate is a discrete mixture of normal distributions where the components' means are Gaussian process realizations observed at a spatial location. An extension of this model allowing different surface selection at different sites was proposed by Duan, Guindani, and Gelfand (2007). In turn, Griffin and Steel (2006) proposed a spatial DP model that permutes the random variables building the weights in the stick-breaking representation, allowing the occurrence of the stick-breaking atoms to be more or less likely in different regions of the spatial domain. Further, Reich and Fuentes (2007), motivated by the need to analyze hurricane surface wind fields, developed a spatial stick-breaking prior where the weights are spatially correlated. Related proposals include Petrone, Guindani and Gelfand (2009) and Rodríguez, Dunson, and Gelfand (2010), both developing spatial DP's where the stick-breaking weights follow a copula representation. Very recently, Zhou, Hanson and Knapp (2015) considered a spatial model where the marginal distributions follow the linear DDP of De Iorio et al. (2009), but a copula induces dependence for georeferenced data. The local DP (Chung and Dunson, 2011), developed to accommodate predictor dependent weights in a DDP with identical margins, offers an approach to the localized spatial 'sharing' of atoms within neighborhoods of fixed size that could be extended to the spatial setting. Additionally, Fuentes and Reich (2013) generalize the models of Reich and Fuentes (2007) and Dunson and Park (2008) to the multivariate spatial setting with non-separable and non-stationary covariance functions. A related approach, although not relying on the stick-breaking representation, was developed by Jo et al. (2017), who considered spatial conditional autoregressive (CAR) species sampling models. In contrast, the frequentist literature on spatial density estimation is very scarce, with an exception being the spatially weighted kernel density estimator proposed by Brunsdon and Charlton (2002) (Section 8.4, pp. 202–203).

All of the above methods, as already mentioned, rely on discrete mixtures of smooth kernels; in fact, each is a particular mixture of normal distributions with some subset of model parameters changing smoothly in space. To the best of our knowledge, the only approach to spatial density estimation that does not rely on mixtures is the very recent Polya tree approach of Tansey et al. (2017). This approach follows Zhao and Hanson (2011) by taking the conditional probabilities that define the Polya tree to have a spatial structure. Whereas Zhao and Hanson (2011) consider multiple logistic-transformed independent CAR priors for the Polya tree conditional probabilities over a lattice (i.e., areal data), in Tansey et al. (2017) the logistic-transformed Polya tree conditional probabilities from adjacent spatial locations are shrunk toward each other via a graph-fused LASSO prior. This latter approach is especially fast and easy to compute when spatial locations lie on a rectangular grid. However, since the approach is not marginalized, density estimates at every spatial location need to be computed, and therefore the method is not immediately amenable to multivariate outcomes. Furthermore, an important drawback of Tansey et al. (2017) is that the fitting algorithm

requires spatial locations to fall on a rectangular grid, something that rarely happens with typical observational data, data arising from irregularly-placed monitoring stations, et cetera.

Our proposed estimator is built upon a modification of the predictive density from a marginalized Polya tree to accommodate localized behavior. The modification is readily implemented in existing Markov chain Monte Carlo (MCMC) schemes for models using marginalized Polya trees (e.g., Hanson, 2006). Because we rely on marginalization, the method is fast, even for multivariate outcomes. Additionally, and unlike DP priors based methods, our method does not rely on mixtures; in fact, a nice feature of Polya trees is that they can be centered at a parametric family (e.g., a normal distribution). In a similar fashion to the work of Dunson (2007), Dunson, Pillai, and Park (2007), and Dunson and Park (2008), observations are weighted according to a distance measure. However, unlike these approaches, a key property of our model is that for extreme covariate values, the estimator essentially follows the parametric family centering the Polya tree. That is, in spatial regions where data are sparse, the estimate is smoothed towards the parametric family, whereas areas that are data-heavy provide a more data-driven estimate. Also, our estimator can handle spatial locations (e.g., longitude and latitude), or covariates, or a mixture of spatial location and covariates. Additional contributions of our work include a refinement to accommodate arbitrarily censored data and a test for whether the density changes across space (and/or covariates). It is worth mentioning that although we are mainly interested in density regression, the methods developed here can be used to model the error in a general regression setup.

The remainder of the paper is organized as follows. In Section 2 we present our model, associated MCMC scheme, a generalization to censored data, and a permutation test. Section 3 provides several applications of the methods to real data. Concluding remarks are provided in Section 4.

2 The predictive model

The Polya tree, and other partition models, lead to a beautifully simple updating rule. A family of densities $\{G_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ is assumed to approximately hold, and $\mathbb{R}$ is broken up into regions of equal probability $\frac{1}{2^j}$ under $G_{\boldsymbol{\theta}}$ at level j. As data are collected, the proportion falling into a region is compared to what is expected under $G_{\boldsymbol{\theta}}$; if this proportion is higher than expected under $G_{\boldsymbol{\theta}}$, then the region is assigned higher predictive probability, and vice versa. However, the amount of the increase or decrease relative to $G_{\boldsymbol{\theta}}$ is attenuated through a smoothing parameter c, that signifies how much confidence one has in the family $G_{\boldsymbol{\theta}}$ to begin with. This simple updating rule requires only counting the numbers of observations falling into the regions. This is now developed formally.

Initially assume data follow a univariate Polya tree centered at a normal distribution

$$y_1, \ldots, y_n | G \overset{\text{iid}}{\sim} G, \quad G \sim PT_J(c, N(\mu, \sigma^2)),$$

truncated to level J (Hanson, 2006). The predictive density for an observation y_i given the previous values $\mathbf{y}_{1:i-1} = (y_1, y_2, \ldots, y_{i-1})$ is given by

$$p(y_i|\mathbf{y}_{1:i-1}, c, \boldsymbol{\theta}) = \phi(y_i|\boldsymbol{\theta}) \prod_{j=1}^{J} \frac{cj^2 + \sum_{k=1}^{i-1} I\{\lceil 2^j \Phi\{\frac{y_i-\mu}{\sigma}\}\rceil = \lceil 2^j \Phi\{\frac{y_k-\mu}{\sigma}\}\rceil\}}{cj^2 + \frac{1}{2}\sum_{k=1}^{i-1} I\{\lceil 2^{j-1} \Phi\{\frac{y_i-\mu}{\sigma}\}\rceil = \lceil 2^{j-1} \Phi\{\frac{y_k-\mu}{\sigma}\}\rceil\}}, \quad (1)$$

where $\phi(y|\boldsymbol{\theta})$ is the density function for a $N(\mu, \sigma^2)$ random variable, $\boldsymbol{\theta} = (\mu, \log\sigma)$, $I\{A\}$ is the usual indicator function for event A, $\Phi(\cdot)$ is the cumulative distribution function of $N(0,1)$, $\lceil\cdot\rceil$ is the ceiling function, and c is a precision parameter controlling how closely G follows the parametric centering distribution $N(\mu, \sigma^2)$ in terms of L_1 distance (Hanson, 2006). Large values of c (e.g., 100 or 1000) lead to a strong belief that y_is are closely iid from $N(\mu, \sigma^2)$. On the other hand, smaller values of c (e.g., 0.01 or 0.1) allow more pronounced deviations of G from $N(\mu, \sigma^2)$. Therefore, the predictive density in (1) can change dramatically with c. To mitigate the effect of c on the posterior inference, Hanson (2006) suggests a gamma prior on c which will also be used for our proposal in equation (2) below.

Now consider spatial data where the observation y_i is observed at spatial location $\mathbf{x}_i$. The full data are then $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$. Note that $\mathbf{x}_i$ can simply be spatial location, e.g. longitude and latitude, or covariates, or a mixture of spatial location and covariates. In the Polya tree an observation y_k $(k < i)$ contributes the same weight to the predictive density (1) of y_i, regardless of how close corresponding spatial locations $\mathbf{x}_i$ and $\mathbf{x}_k$ are. If we replace the indicator functions in (1) by a distance measure $d_\psi(\mathbf{x}_i, \mathbf{x}_k)$ only giving a 'whole observation' when $\mathbf{x}_i = \mathbf{x}_k$ and some fraction of unity that is a function of the distance between $\mathbf{x}_i$ and $\mathbf{x}_k$ otherwise, we obtain a predictive process with a tailfree flavor, but the additional flexibility to be able to adapt locally:

$$p(y_i|\mathbf{y}_{1:i-1}, c, \boldsymbol{\theta}, \psi) = \phi(y_i|\boldsymbol{\theta}) \prod_{j=1}^{J} \frac{cj^2 + \sum_{k=1}^{i-1} I\{\lceil 2^j \Phi\{\frac{y_i-\mu}{\sigma}\}\rceil = \lceil 2^j \Phi\{\frac{y_k-\mu}{\sigma}\}\rceil\} d_\psi(\mathbf{x}_i, \mathbf{x}_k)}{cj^2 + \frac{1}{2}\sum_{k=1}^{i-1} I\{\lceil 2^{j-1} \Phi\{\frac{y_i-\mu}{\sigma}\}\rceil = \lceil 2^{j-1} \Phi\{\frac{y_k-\mu}{\sigma}\}\rceil\} d_\psi(\mathbf{x}_i, \mathbf{x}_k)} \quad (2)$$

The distance measure used herein is a function of the sample Mahalanobis distance $d_\psi(\mathbf{x}_i, \mathbf{x}_k) = \exp\{-\psi(\mathbf{x}_i - \mathbf{x}_k)'\mathbf{S}^{-1}(\mathbf{x}_i - \mathbf{x}_k)\}$, defining an effective window in which data can affect the predictive density. When $\psi = 0$ the prediction rule from a Polya tree with exchangeable observations is obtained as $d_\psi(\mathbf{x}_i, \mathbf{x}_k) = 1$ for all i and k. When $d_\psi(\mathbf{x}_i, \mathbf{x}_k) \approx 0$, it is essentially as if y_k is not in the sample. Note for $\mathbf{x}_i$ very far from the sample mean $\bar{\mathbf{x}}$, the distances will all essentially be zero and the parametric model is obtained. As a consequence, in regions that are data scarce the estimator will essentially follow the parametric $N(\mu, \sigma^2)$ centering the Polya tree.

The Mahalanobis distance gives the commonly-used spatial Gaussian correlation function when Euclidean coordinates are independently observed. Furthermore the Mahalanobis distance is anisotropic, allowing for quite different "x" and "y" scales if present in data, as well as correlated spatial coordinates. Distance measures such as this are also used in "local Dirichlet process" approaches, e.g. Chung and Dunson

(2011) as well as earlier versions for densities that change smoothly with covariates, e.g. Dunson (2007), although not incorporating correlation among variables. The accommodation of correlation is important because it essentially obviates concerns about multicollinearity, as well as mitigates a need for variable selection. Superfluous dimensions of $\mathbf{x}_i$ are handled naturally within the Mahalanobis distance wherein the distances between highly positively correlated variables are much less than the same distances between uncorrelated or negatively correlated variables. We reiterate that the "locations" $\mathbf{x}_i$ can be spatial locations, covariates, or mixtures of spatial locations and covariates.

The joint density for all observations is given by the Markov expansion

$$p(y_1, \ldots, y_n | c, \boldsymbol{\theta}, \psi) = \prod_{i=1}^{n} p(y_i | \mathbf{y}_{1:i-1}, c, \boldsymbol{\theta}, \psi), \tag{3}$$

where $p(y_1|c, \boldsymbol{\theta}, \psi) = \phi(y_1|\boldsymbol{\theta})$. This defines a valid joint probability model; however, $p(y_1, \ldots, y_n|c, \boldsymbol{\theta}, \psi)$ is not invariant to the order in which pairs $(\mathbf{x}_i, y_i)$ enter into the model. That is, $p(y_1, y_2, y_3|c, \boldsymbol{\theta}, \psi)$ is not necessarily equal to $p(y_2, y_3, y_1|c, \boldsymbol{\theta}, \psi)$. This has ramifications in terms of drawing inferences for $(c, \boldsymbol{\theta}, \psi|\mathbf{y}_{1:n})$.

The posterior

$$p(c, \boldsymbol{\theta}, \psi | \mathbf{y}_{1:n}) \propto p(y_1, \ldots, y_n | c, \boldsymbol{\theta}, \psi) p(c, \boldsymbol{\theta}, \psi)$$

depends on the order $y_1, \ldots, y_n$. To make the posterior invariant to the ordering of the observed data, we propose to instead use the permutation density

$$p(c, \boldsymbol{\theta}, \psi | \mathbf{y}_{1:n}) \propto p(c, \boldsymbol{\theta}, \psi) \frac{1}{n!} \sum_{(i_1, \ldots, i_n) \in \mathcal{P}} p(y_{i_1}, \ldots, y_{i_n} | c, \boldsymbol{\theta}, \psi),$$

where $\mathcal{P}$ are all permutations of $\{1, \ldots, n\}$. Dahl, Day, and Tsai (2017) consider a related scenario in defining a partition distribution indexed by a permutation. Computation of all permutations is not feasible so Dahl et al. (2017) place a uniform prior distribution over all permutations, allowing the MCMC to numerically perform the marginalization. In our setting there is no reason to favor one permutation over another so equal weight is placed on all partitions by the *model* rather than through only the prior. Thus the partition is not "updated" during MCMC according to a Metropolis-Hastings step; rather a different permutation is forced at each MCMC iteration.

Although the permutation density avoids order dependence, this dependence is empirically observed to be quite weak. Simply using the data order as observed changes posterior inference only slightly from that obtained through the permutation density. Similarly, Newton and Zhang (1999) noted in a Bayesian nonparametric setting involving the Dirichlet process with censored data, where exchangeability is lost, the order of updating affects the predictive distribution negligibly.

We proceed to develop a reasonable prior for ψ. With probability q, assume that the density is not spatially-varying, i.e. $P(\psi = 0) = q$; setting $q = 0.5$ gives the posterior odds as the Bayes factor so we consider that here. For spatially-varying densities assume $\psi|\psi > 0 \sim \Gamma(a_\psi, b_\psi)$, where $\Gamma(a, b)$ denotes a gamma distribution with mean a/b. We set $a_\psi = 2$ and $b_\psi = (a_\psi - 1)/\psi_0$ so that the prior of $\psi|\psi > 0$ has mode at ψ_0, where ψ_0 satisfies $\exp\{-\psi_0 \max_{1\le i\le n}(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})\} = 0.001$. Thus the final prior on ψ is the mixture

$$\psi \sim q\ \delta_0 + (1 - q)\ \Gamma(a_\psi, b_\psi),$$

where δ_x is Dirac measure at x and the default is $q = \frac{1}{2}$. A Bayes factor for comparing the spatially varying model to the exchangeable model is given by

$$BF = \frac{P(\psi > 0|\mathbf{y}_{1:n})/P(\psi = 0|\mathbf{y}_{1:n})}{P(\psi > 0)/P(\psi = 0)} = \frac{q\ P(\psi > 0|\mathbf{y}_{1:n})}{(1 - q)\ P(\psi = 0|\mathbf{y}_{1:n})}.$$

Two options are taken for $(\mu, \log\sigma)$. The default follows Hanson, Branscum, and Gardner (2008) and Chen and Hanson (2014) by simply fixing them at their maximum likelihood estimates (MLEs) under the parametric $N(\mu, \sigma^2)$ model as $c \to \infty$; for uncensored data these are of course the sample moments $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n y_i$, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \hat{\mu})^2$. When data are censored, the MLEs are not available in closed-form, but are readily computed by most statistical software packages, e.g. `survreg` in R using `family="gaussian"`. The second option is to take a bivariate normal prior $N_2(\boldsymbol{\theta}_0, \mathbf{V}_0)$ on $\boldsymbol{\theta} = (\mu, \log\sigma)$ with $\boldsymbol{\theta}_0 = (\hat{\mu}, \log\hat{\sigma})$ and $\mathbf{V}_0$ being the estimated asymptotic "plug-in" covariance for the MLE. The only remaining parameters is c; $c \sim \Gamma(a_c, b_c)$ is assumed with the default $c \sim \Gamma(5, 1)$.

The function `SpatDensReg` is developed within the R package `spBayesSurv` (Zhou and Hanson 2018) for the implementation. The usage syntax is

```
SpatDensReg(formula, data, na.action, prior=NULL, state=NULL,
            mcmc=list(nburn=3000, nsave=2000, nskip=0, ndisplay=500),
            permutation=TRUE, fix.theta=TRUE)
```

Here `formula` is a formula expression with the response returned by the `Surv` function in the `survival` package. It supports right-censoring, left-censoring, interval-censoring, and mixtures of them. For survival data, the input response should be log survival times. The argument `prior` is a list giving the prior information. The list includes the following elements.

`prior` element	`maxL`	`a0`	`b0`	`theta0`	`V0`	`phia0`	`phib0`	`phiq0`
corresponding symbol	L	a_c	b_c	$\boldsymbol{\theta}_0$	$\mathbf{V}_0$	a_ψ	b_ψ	q

The argument `permutation` is a logical flag to indicate whether a random data permutation will be implemented in the beginning of each iterate; the default is `TRUE`. The argument `fix.theta` is a logical flag to indicate whether the parameters $\boldsymbol{\theta}$ are fixed; the default is `TRUE` indicating that they are fixed at $(\hat{\mu}, \log\hat{\sigma})$.

2.1 Markov chain Monte Carlo

The MCMC algorithm uses blockwise adaptive MCMC (Haario, Saksman, and Tamminen, 2001). The blocks are $(\mu, \log\sigma)$ when the option `fix.theta=FALSE` is chosen, c, and ψ; ψ requires some special care, detailed below. The permutation density can be used by fixing `permutation=TRUE` which is the default.

For the mixture prior $\psi \sim q\delta_0 + (1-q)\Gamma(a_\psi, b_\psi)$ we need to first conditionally sample whether $\psi = 0$ or not. Bayes rule gives

$$P(\psi = 0|\mathbf{y}_{1:n}, c, \boldsymbol{\theta}) = \frac{q\ p(\mathbf{y}_{1:n}|c, \boldsymbol{\theta}, \psi = 0)}{q\ p(\mathbf{y}_{1:n}|c, \boldsymbol{\theta}, \psi = 0) + (1-q)\int_0^\infty p(\mathbf{y}_{1:n}|c, \boldsymbol{\theta}, \psi)\gamma(\psi; a_\psi, b_\psi)d\psi},$$

where $\gamma(\cdot; a, b)$ refers to the density of $\Gamma(a, b)$. Set $\Gamma(\cdot; a, b)$ to be the cumulative distribution function of $\Gamma(a, b)$. The integral can be approximated by a Riemann sum

$$\int_0^\infty p(\mathbf{y}_{1:n}|c, \boldsymbol{\theta}, \psi)\gamma(\psi; a_\psi, b_\psi)d\psi \approx \sum_{k=1}^{K} \frac{1}{\psi_k - \psi_{k-1}} p(\mathbf{y}_{1:n}|c, \boldsymbol{\theta}, \psi_k)\gamma(\psi_k; a_\psi, b_\psi),$$

where $\psi_k = \Gamma^{-1}(\frac{k}{K+1}; a_\psi, b_\psi)$ for $k = 0, 1, \ldots, K$, e.g. $K = 20$. If $\psi = 0$ is sampled, we are done, otherwise we need to sample $\psi|\psi > 0$ using usual adaptive M-H. When computing the adaptive variance, only those sample values of ψ that are positive are included. An option forcing $\psi > 0$ only $(q = 0)$ is also available which speeds up the MCMC considerably but disallows the computation of a Bayes factor to test whether spatial location and/or covariates affect the distribution of the response.

Given the posterior sample $\{\psi^{(j)}, j = 1, \ldots, M\}$, the Bayes factor for the spatial model vs. exchangeable model is simply $[\frac{1-\bar{q}}{\bar{q}}]/[\frac{1-q}{q}]$ where $\bar{q} = \frac{1}{M}\sum_{j=1}^{M} I\{\psi^{(j)} = 0\}$.

2.2 Censored data

Censored observations are readily sampled from Metropolis-Hastings proposals based on the underlying centering distribution. Define

$$q(y_i|\mathbf{y}_{-i}, c, \boldsymbol{\theta}, \psi) = \prod_{j=1}^{J} \frac{cj^2 + \sum_{k\neq i} I\{\lceil 2^j\Phi\{\frac{y_i-\mu}{\sigma}\}\rceil = \lceil 2^j\Phi\{\frac{y_k-\mu}{\sigma}\}\rceil\}d_\psi(\mathbf{x}_i, \mathbf{x}_k)}{cj^2 + \frac{1}{2}\sum_{k\neq i} I\{\lceil 2^{j-1}\Phi\{\frac{y_i-\mu}{\sigma}\}\rceil = \lceil 2^{j-1}\Phi\{\frac{y_k-\mu}{\sigma}\}\rceil\}d_\psi(\mathbf{x}_i, \mathbf{x}_k)},$$

where $\mathbf{y}_{-i}$ is $\mathbf{y}_{1:n}$ with the ith observation removed.

Let $\mathcal{C} = \{i : \delta_i = 0\}$ be the indices of censored observations where $\delta_i = 0$ if y_i is only known to lie in the interval $y_i \in (a_i, b_i)$, $a_i < b_i$, and $\delta_i = 1$ if y_i is observed exactly. The latent values of $\mathcal{Y}_c = \{y_i : i \in \mathcal{C}\}$ are updated via MCMC along with the model parameters $\boldsymbol{\theta}$, c, and ψ. If $i \in \mathcal{C}$ propose $y_i^* \sim N(\mu, \sigma^2)$ truncated to (a_i, b_i) and accept with probability

$$1 \wedge \frac{q(y_i^*|\mathbf{y}_{-i}, c, \boldsymbol{\theta}, \psi)}{q(y_i|\mathbf{y}_{-i}, c, \boldsymbol{\theta}, \psi)},$$

otherwise leave y_i at it's current value.

2.3 Direct estimation and a permutation test p-value

For uncensored data, estimation can be sped up substantially by avoiding MCMC entirely and simply using maximum a posteriori (MAP) estimates coupled with an empirical Bayes approach to fixing the centering distribution. Dunson (2007) considered a somewhat related approach in a covariate-weighted Dirichlet process mixture. Chen and Hanson (2014) consider a hybrid approach for uncensored data by setting $\boldsymbol{\theta} = (\hat{\mu}, \log \hat{\sigma})$, the MLE's under normality, and maximizing (3) with $\psi = 0$ over a grid of c values $c_i = \exp\{\frac{14}{19}(i-1) - 7\}$ for $i = 1, \ldots, 20$. The spatial version (3) allowing for $\psi > 0$ can similarly be maximized over a lattice of (c_i, ψ_j) values. From considerations in Section 2.1, quantiles of the prior $\psi_j = \Gamma^{-1}(\frac{j}{11}; a_\psi, b_\psi)$ for $i = 1, \ldots, 10$ are reasonable; call these values $\mathcal{S}$. Similarly $c_i \in \mathcal{C} = \{0.001, 0.01, 0.1, 0.5, 1, 5, 10, 50, 100, 1000\}$ could be used giving 100 values of $\{(c_i, \psi_j)\}$ to compute and maximize (3) over.

The Bayes factor described in Sections 2.1 and 2.2 tests the hypothesis $H_0 : \psi > 0$ relative to $H_0 : \psi = 0$ via MCMC using the priors described in these sections. A "maximized Bayes factor" from direct estimation is given by

$$BF = \frac{\max_{(c,\psi) \in \mathcal{C} \times \mathcal{S}} p(c, \hat{\boldsymbol{\theta}}, \psi | \mathbf{y}_{1:n})}{\max_{c \in \mathcal{C}} p(c, \hat{\boldsymbol{\theta}}, 0 | \mathbf{y}_{1:n})}. \tag{4}$$

This Bayes factor gives the "most evidence" in favor of the spatially-varying model and akin to a likelihood ratio test, albeit with added prior information and a plug-in estimate for $\boldsymbol{\theta}$. Consider the null $H_0 : \mathbf{x} \in \mathcal{X}$ is independent of $y \in \mathcal{Y}$. Under this null we can repeatedly take random, uniformly distributed permutations $(i_1, \ldots, i_n) \in \mathcal{P}$, form "data" $\{(\mathbf{x}_j, y_{i_j})\}_{j=1}^n$, and compute Bayes factors from (4). The proportion of these larger than the one based on the original data is a permutation test p-value (Fisher, 1935) for testing association between the response and spatial location (and/or covariates).

The function `BF.SpatDensReg` is developed within the R package `spBayesSurv` to obtain the BF in (4) and the permutation test p-value. The usage syntax is

```
BF.SpatDensReg(y, X, prior = NULL, nperm = 100, c_seq = NULL,
               phi_seq = NULL)
```

Here `y` is a vector of uncensored responses, rows of `X` are spatial locations and/or covariates, `prior` is the same as the one used in `SpatDensReg`, `nperm` is an integer giving the total number of permutations, `c_seq` is an vector giving grid values for c, and `phi_seq` is a vector giving grid values for ψ. To illustrate the use of this method, we generate the data as follows: $y_i \overset{ind}{\sim} N(\beta x_i, 0.2^2)$, $x_i \overset{iid}{\sim} \text{Beta}(0.3, 0.3)$, $i = 1, \ldots, 300$, where $\beta = 0.01, 0.05, 0.1, 0.5, 1$. The following R code is used to obtain these BFs and p-values. As expected, we see that as β increases, so does the Bayes factor while the p-value is approaches zero.

```
library(spBayesSurv)
```

```
set.seed(2017)
beta = c(0.01, 0.05, 0.1, 0.5, 1);
BFs = rep(NA, length(beta));
Pvalues = rep(NA, length(beta));
for(sim in 1:length(beta)){
print(sim);
## Generate data
n = 300;
x = rbeta(n, 0.3, .3)
y = rep(0, n);
uu = runif(n);
for(i in 1:n){
y[i] = rnorm(1, beta[sim]*x[i], .2);
}
prior = list(maxL=6);
res1 = BF.SpatDensReg(y, x, prior=prior, nperm=500);
BFs[sim] = res1$BF;
Pvalues[sim] = res1$pvalue;
}
###### Outputs:
> BFs
   beta=0.01    beta=0.05     beta=0.1     beta=0.5       beta=1
4.624762e-01 3.283394e+00 8.828178e+03 1.290706e+30 4.403196e+83
> Pvalues
beta=0.01 beta=0.05  beta=0.1  beta=0.5    beta=1
    0.974     0.922     0.000     0.000     0.000
```

3 Examples

3.1 IgG distribution evolving with age

Jara and Hanson (2011) and Schörgendorfer and Branscum (2013) considered serum immunoglobulin G (IgG) concentrations from $n = 298$ children aged 6 months to 6 years old. Like these authors we consider the log-transformation of the data y_i; the log-IgG values are plotted versus age in Figure 1(f). We consider the spatially smoothed Polya tree for estimating the log-IgG density as smoothly varying function of age. Unlike previous authors we rely on only the Polya tree and do not explicitly model an IgG trend via fractional polynomials.

The following R code is used to fit the proposed model with $J = 4$ and the default prior settings in Section 2 except for the option `fix.theta=FALSE` which provides much smoother posterior density estimates.

```
#needed packages
library(survival)
library(spBayesSurv)
library(coda)
library(DPpackage)

#data management
data(igg); d = igg; n = nrow(d);
d$logIgG = log(d$igg)

#fitting the model
nburn=20000; nsave=5000; nskip=9;
mcmc=list(nburn=nburn, nsave=nsave, nskip=nskip, ndisplay=500);
prior = list(maxL=4, phiq0=0.5);
res1 = SpatDensReg(formula = Surv(logIgG)~age, data=d, prior=prior,
   mcmc=mcmc, permutation = TRUE, fix.theta=FALSE);

#output from summary
summary(fit) # most output removed to save space
Posterior inference of centering distribution parameters
(Adaptive M-H acceptance rate: 0.1054):
             Mean      Median    Std. Dev.  95%CI-Low  95%CI-Upp
location     1.47804   1.49332   0.05165    1.33244    1.53099
log(scale)  -0.70716  -0.70207   0.04997   -0.81437   -0.62347

Posterior inference of precision parameter
(Adaptive M-H acceptance rate: 0.35896):
       Mean    Median  Std. Dev.  95%CI-Low  95%CI-Upp
alpha  0.6967  0.6368  0.3025     0.2866     1.4248

Posterior inference of distance function range phi
(Adaptive M-H acceptance rate: 0.38898):
       Mean   Median  Std. Dev.  95%CI-Low  95%CI-Upp
range  3.527  3.326   1.212      1.805      6.484

Bayes Factor for the spatial model vs. the exchangeable model: Inf
Number of subjects: n=298
```

The traceplots for $\boldsymbol{\theta}$, ψ and c mixed very well (not shown). The Bayes factor for testing association between age and log-IgG is ∞ indicating a decisive evidence of dependency. Figures 1 presents the posterior mean and 95% pointwise credible interval of the log-IgG density at five different ages. These fitted densities are similar to those obtained by Jara and Hanson (2011). The following R code can be used to provide these plots.

```
ygrid = seq(min(d$logIgG), max(d$logIgG), length.out = 200);
xpred = data.frame(age=c(11, 25, 38, 52, 65, 79)/12);
estimates=plot(res1, xnewdata=xpred, ygrid=ygrid);
for(i in 1:nrow(xpred)){
pdf(file =paste("IgG-densities-age", xpred[i,]*12, "-bf.pdf", sep=""),
paper="special", width=8, height=6)
par(cex=1.5,mar=c(4.1,4.1,2,1),cex.lab=1.4,cex.axis=1.1)
plot(estimates$ygrid, estimates$fhat[,i], "l",
main=paste("IgG data, age=", xpred[i,]*12, "months", sep=""),
xlab="log IgG", ylab="density",
ylim=c(0,1.5), xlim=c(0, 2.5), lty=1, lwd=3);
lines(estimates$ygrid, estimates$fhatup[,i], lty=2, lwd=2);
lines(estimates$ygrid, estimates$fhatlow[,i], lty=2, lwd=2);
dev.off()
}
pdf(file ="IgG-scatter.pdf", paper="special", width=8, height=6)
par(cex=1.5,mar=c(4.1,4.1,2,1),cex.lab=1.4,cex.axis=1.1)
plot(d$age*12, d$logIgG, main="IgG data",
xlab="age (months)", ylab="log IgG")
for(i in 1:5){
points(xpred[i,]*12, -0.2, pch = 16, cex=1.3,
col = "red", las = 1,xpd = TRUE)
text(xpred[i,]*12, 0, paste("", xpred[i,]*12, sep=""),
col = "red", adj = c(-0.1, .5))
}
dev.off()
```

3.2 Time to infection in amphibian populations

Spatial data on the number of years from discovery to the time-to-arrival of the fungus *Batrachochytrium dendrobatidis* (Bd) in mountain yellow-legged frog populations throughout Sequoia-Kings Canyon National Park was considered by Zhou, Hanson, and Knapp (2015). Once infected, the Bd fungus can wipe out a frog population in a few weeks, and it is of interest to determine the distribution of time-to-infection and how it varies spatially. The data consist of $n = 309$ frog populations (Figure 2(f)) initially discovered during park-wide surveys conducted from 1997 to 2002, and then resurveyed regularly through 2011. The observed event time is calculated as the number of years from the initial survey to either Bd arrival (time actually observed) or the last resurvey (right censored). By the end of the study, about 11% of the frog populations remained Bd-negative (right censored), and the rest of populations are interval censored.

We fit the spatially smoothed Polya tree with the same settings as Section 3.1. The Bayes factor for testing spatial variation of time-to-Bd is estiamted to be ∞, strong

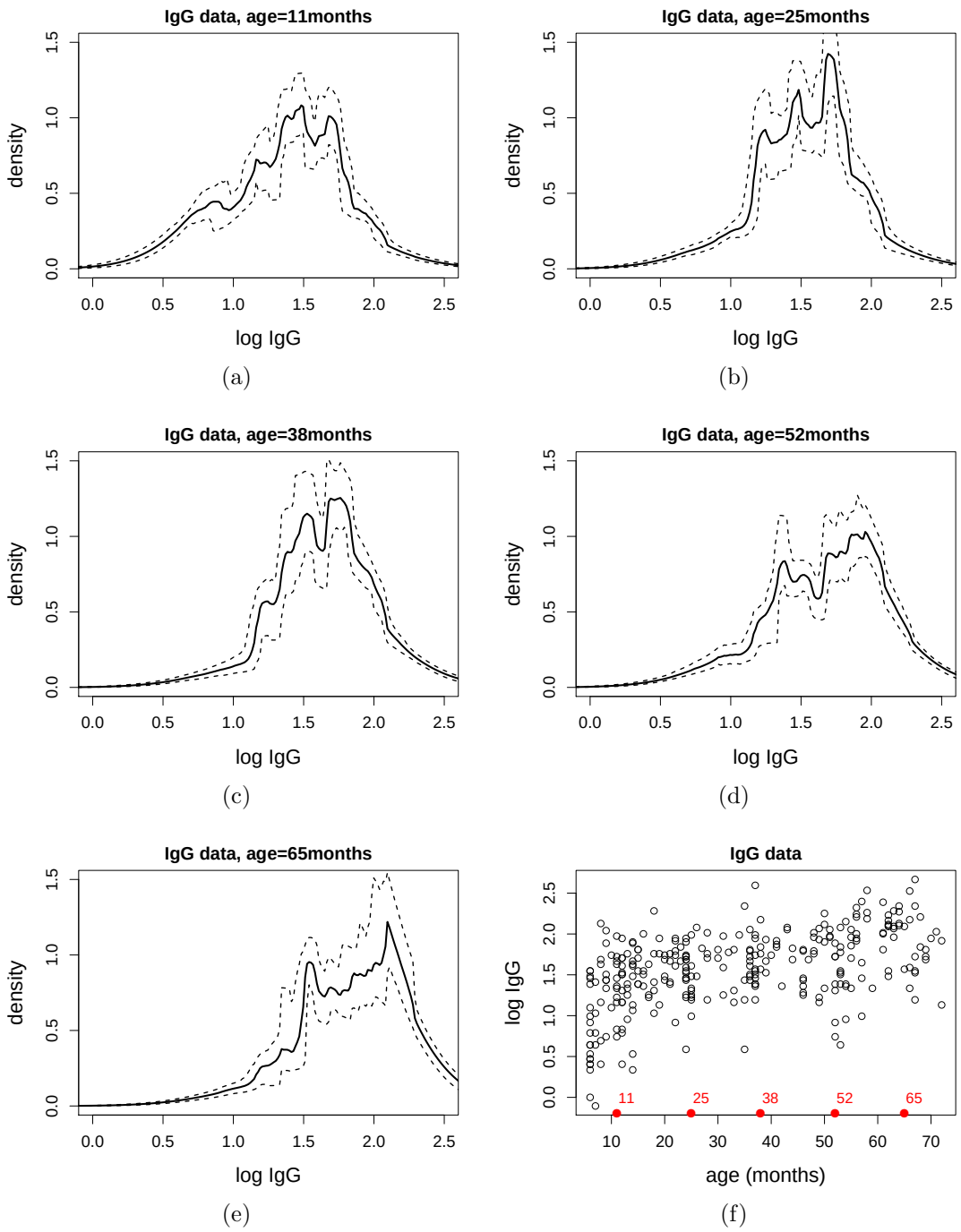

Figure 1: IgG data. Panels (a)-(e) shows the posterior mean (solid) and 95% pointwise credible interval (dashed) of the density of log IgG at five different ages. Panel (f) shows the five age points and the scatter plot of the data.

evidence that the time-to-Bd distribution spatially varies. Figure 2 shows the posterior mean and 95% pointwise credible intervals of the log time-to-Bd density at five different locations (marked in Figure 2(f)). The distribution of log time-to-Bd at location 5 has two modes which can also be seen from the predictive log times-to-Bd around this location.

3.3 Simulated data

We generate two datasets from the following two scenarios, respectively.

1. $y_i \overset{ind}{\sim} N(x_i, 0.2^2)$, $x_i \overset{iid}{\sim} \text{Beta}(0.3, 0.3)$, $i = 1, \ldots, 300$,
2. $y_i \overset{iid}{\sim} 0.5N(0.5, 0.2^2) + 0.5N(1, 0.3^2)$, $x_i \overset{iid}{\sim} \text{Beta}(0.3, 0.3)$, $i = 1, \ldots, 300$.

Here we expect a large BF value for scenario 1 and a BF less than 1 for scenario 2. The censoring times are generated from Uniform$(0.5, 2)$ so that the censoring rate is 0.13 under scenario 1 and 0.21 under scenario 2.

The spatially smoothed Polya tree is fit with $J = 6$ and the same prior settings as Section 3.1. We retain 5,000 scans thinned from 50,000 after a burn-in period of 20,000 iterations. The BF factor for scenario 1 is ∞ as expected, while it is 0.01 under scenario 2. Figures 3 and 4 present the posterior mean and 95% pointwise credible interval of the conditional density of y at three different x values under each scenario. The results demonstrate that the proposed model can capture the conditional densities quite well *without any spatial trend component*, although the estimates are a bit spiky.

To investigate the impact of censoring on our model performance, we use the simulated data 1 (Figure 3(d), uncensored version) again under the following two cases: (i) right-censoring with high censoring rate, and (ii) interval-censoring. For case (i), the censoring times are generated from $N(x_i - 0.2, 0.5^2)$ yielding a 0.67 right-censoring rate. For case (ii), we first generate right-censored times from Uniform$(0.5, 2)$, then transfer uncensored times into internal-censored times using the endpoints $\{0, 0.2, 0.4, \ldots, 1.8, 2\}$, yielding a rate of 0.13 for right-censoring, 0.14 for left-censoring and 0.73 for interval-censoring. The BF factors for both cases are ∞ as expected. The posterior conditional density estimates (Figure 5) are all close to the truth except for the $x = 0$ and $x = 1$ under case (i) for which increasing the sample size can be helpful. In addition, wider credible intervals are also observed as expected. Overall our method still performs reasonably well for right-censored data with high censoring rate and interval-censored data.

4 Conclusion

The prediction rule from a marginalized Polya tree is generalized to spatially smooth densities over spatial regions, weighing data from proximal locations more heavily than remote ones. Although ideas presented are quite simple and easy-to-implement, the

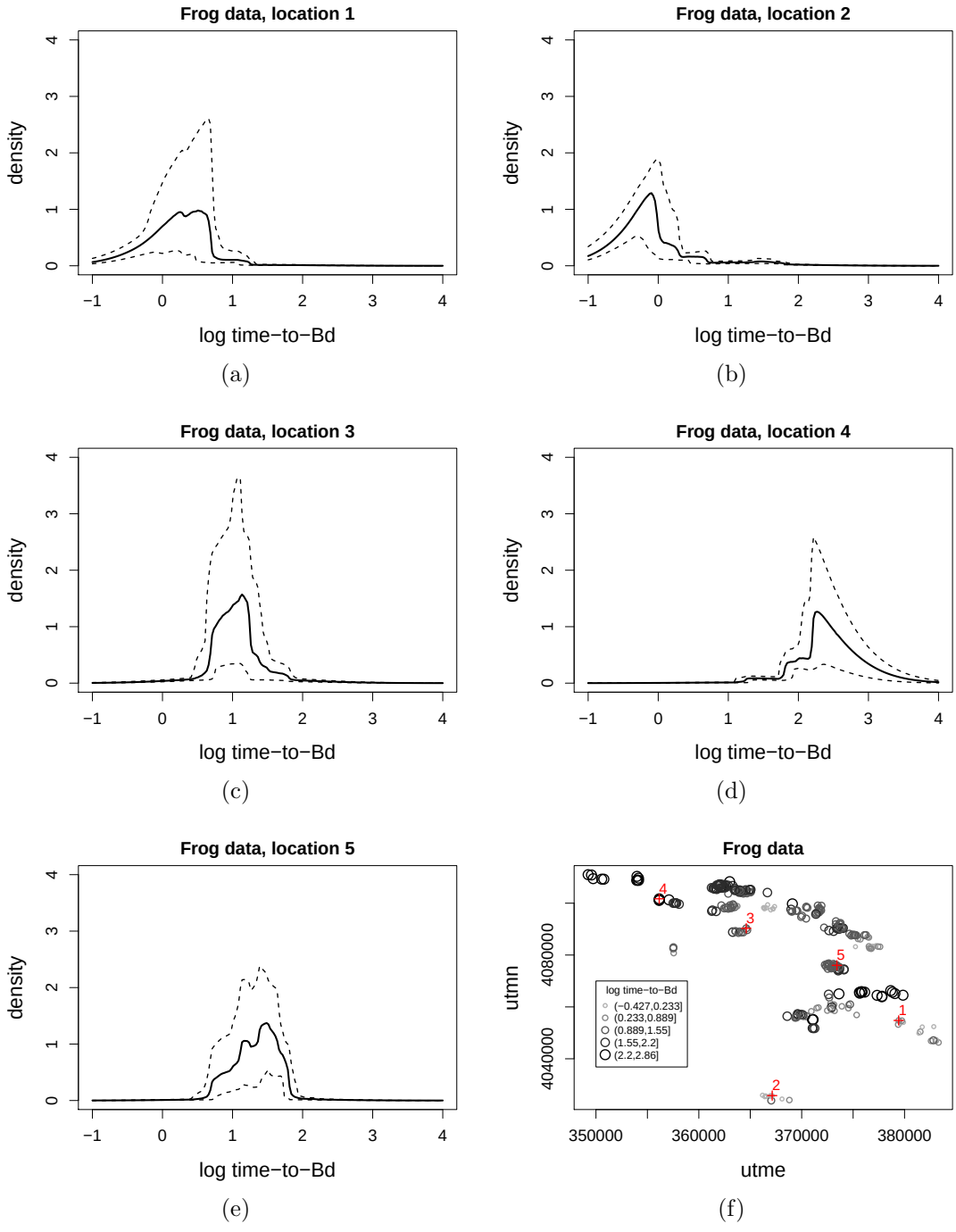

Figure 2: Frog data. Panels (a)-(e) shows the posterior mean (solid) and 95% pointwise credible interval (dashed) of the log time-to-Bd density at five different locations. Panel (f) shows the five considered locations and the data locations with circle size representing the posterior mean of log times-to-Bd.

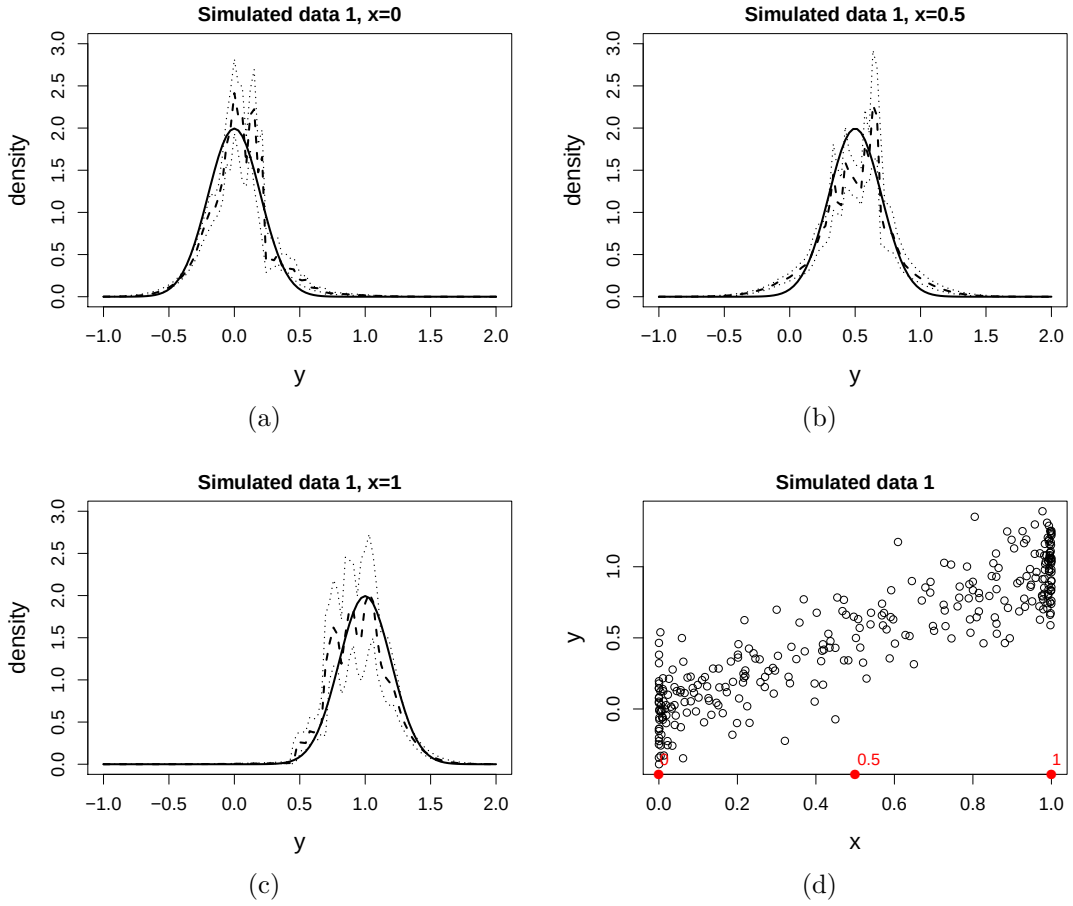

Figure 3: Simulated data 1. Panels (a)-(c) shows the posterior mean (dashed) and 95% pointwise credible interval (dotted) of the conditional density at three different x values; the solid curves are the corresponding true densities. Panel (d) shows the three x points and the scatter plot of the data.

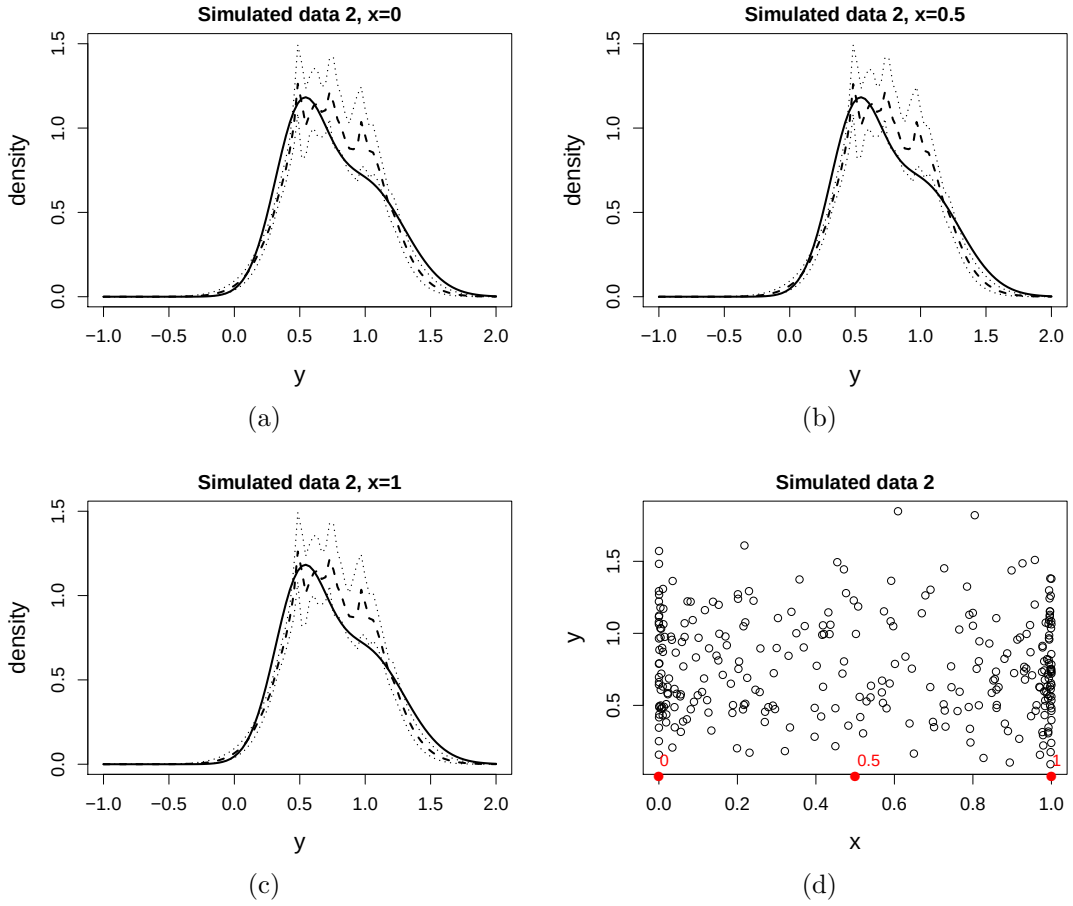

Figure 4: Simulated data 2. Panels (a)-(c) shows the posterior mean (dashed) and 95% pointwise credible interval (dotted) of the conditional density at three different x values; the solid curves are the corresponding true densities. Panel (d) shows the three x points and the scatter plot of the data.

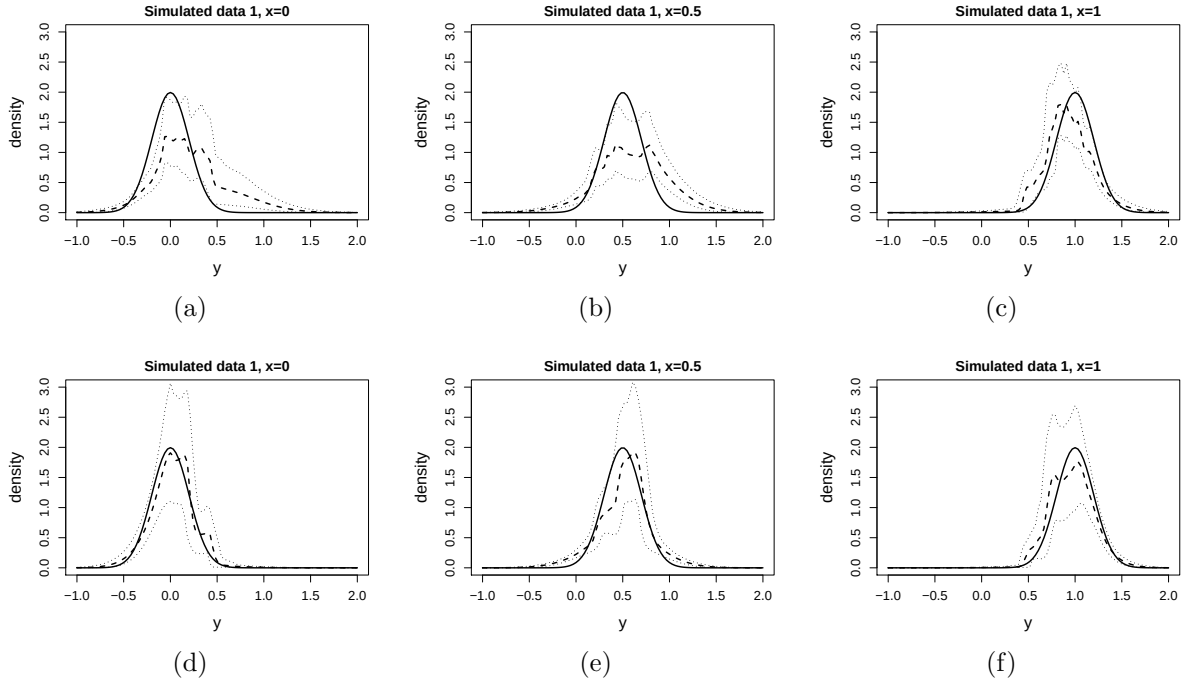

Figure 5: Simulated data 1 with high right-censoring rate (panels a, b, c) and interval-censoring (panels c, d, e). Each panel provides the posterior mean (dashed) and 95% pointwise credible interval (dotted) of the conditional density at three different x values; the solid curves are the corresponding true densities.

approach has several advantages quite distinct from other approaches. First, it is the only method that we are aware of that smooths the density estimate towards a parametric estimate in data-lean portions of space, e.g. a normal density. The method is fairly fast and competitive with methods based on the Dirichlet process. Finally, a freely-available `R` function `SpatDensReg` is available in the `spBayesSurv` package that makes use of compiled C++ to fit the Bayesian model and report the Bayes factor, for arbitrarily censored (or uncensored) data.

As with non-spatial Polya trees, the spatially-smoothed version is easily constrained to be median-zero. Thus median regression with a spatially-weighted error density is possible leading to heteroscedastic accelerated failure time models that retain the interpretability of acceleration factors in terms of the median (e.g. Jara and Hanson, 2011; Zhou, Hanson, and Zhang, 2017). For example, the Polya tree can be shifted and/or stretched via regressions on the centering distribution parameters such as $\mu_{\mathbf{x}} = \mathbf{x}'\boldsymbol{\beta}$ and $\log \sigma_{\mathbf{x}} = \mathbf{x}'\boldsymbol{\tau}$. Similarly, extension to multivariate outcomes is straightforward, including the computation of the Bayes factor for testing spatial dependence; however, obtaining marginal density estimates requires simulating from the Polya tree and using univariate smoothers, e.g. Hanson, Branscum, and Gardner (2008).

One modification of the model as developed that could potentially improve prediction is the use of a spatially-varying centering distribution $\boldsymbol{\theta}_{\mathbf{x}}$ rather than static $\boldsymbol{\theta}$. Spatially weighted $\boldsymbol{\theta}_{\mathbf{x}}$

$$\mu_{\mathbf{x}} = \frac{\sum_{i=1}^{n} d_{\psi}(\mathbf{x}_i, \mathbf{x}) y_i}{\sum_{i=1}^{n} d_{\psi}(\mathbf{x}_i, \mathbf{x})}, \quad \sigma^2_{\mathbf{x}} = \frac{\sum_{i=1}^{n} d_{\psi}(\mathbf{x}_i, \mathbf{x})(y_i - \mu_{\mathbf{x}})^2}{\sum_{i=1}^{n} d_{\psi}(\mathbf{x}_i, \mathbf{x})},$$

are used in the predictive density $p(y|\mathbf{y}_{1:n}, c, \boldsymbol{\theta}_{\mathbf{x}}, \psi)$ so that the location and spread of the centering normal distribution now changes with spatial location. It is unclear, however, how to create a valid likelihood for the remaining parameters (c, ψ) in (2) with spatial $\boldsymbol{\theta}_{\mathbf{x}}$. A possible approach is to simply estimate (c, ψ) via cross-validation methods. This, an exploration of the permutation test for spatial association in Section 2.3, and the median-regression version of the model are topics for future research.

References

Cai, B., Lawson, A., Hossain, M., Choi, J., Kirby, R., and Liu, J. (2013). Bayesian semiparametric model with spatially-temporally varying coefficients selection. *Statistics in Medicine*, **32**, 3670–3685.

Chen, Y., and Hanson, T. (2014). Bayesian nonparametric k-sample tests for censored and uncensored data. *Computational Statistics and Data Analysis*, **71**, 335–346.

Chen, Y., Sun, M., and Hanson, T. Nonparametric multivariate Polya tree EWMA control chart for process changepoint detection. *Statistics and Its Interface*, accepted.

Chung, Y., and Dunson, D. B. (2009). Nonparametric Bayes conditional distribution modeling with variable selection. *Journal of the American Statistical Association* **104**, 1646–1660.

Chung, Y., and Dunson, D.B. (2011). The local Dirichlet process. *Annals of the Institute of Statistical Mathematics* **63**, 59–80.

Cressie, N. (2015). *Statistics for Spatial*, 2nd Edition. Wiley, Hoboken, NJ.

Dahl, D.B., Day, R., and Tsai, J.W. (2017). Random partition distribution indexed by pairwise information. *Journal of the American Statistical Association*, **112**, 721–732.

De Iorio, M., Johnson, W.O., Müller, P., and Rosner, G.L. (2009). Bayesian Nonparametric nonproportional hazards survival modeling, *Biometrics*, **65**, 762–771.

Duan, J., Guindani, M., and Gelfand, A.E. (2007). Generalized spatial Dirichlet process models. *Biometrika* **94**, 809–825.

Dunson, D.B. (2007). Empirical Bayes density regression. *Statistica Sinica*, **17**, 481–504.

Dunson, D.B., and Park, J.-H. (2008). Kernel stick-breaking processes. *Biometrika*, **95**, 307–323.

Dunson, D.B., Pillai, N.S., and Park, J-H. (2007). Bayesian density regression. *Journal of the Royal Statistical Society, Ser. B* **69**, 163–183.

Escobar, M.D. and West M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**, 577–588.

Ferguson T.S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**, 209–230.

Fisher, R. (1935). *The Design of Experiments*. Oliver & Boyd, Edinburgh.

Fotheringham, A.S., Brunsdon, C., and Charlton, M. (2002). *Geographically Weighted Regression: the analysis of spatially varying relationships*, Wiley.

Fuentes, M., and Reich, B. (2013). Multivariate spatial nonparametric modeling via kernel process mixing. *Statistica Sinica*, **23**, 75–97.

Gelfand A.E., Diggle, P.J., Fuentes, M., and Guttorp, P. (eds.) (2010). *Handbook of Spatial Statistics*, Chapman&Hall/CRC Handbooks of Modern Statistical Methods.

Gelfand, A.E., Kottas, A., and Maceachern, S. N. (2005). Bayesian nonparametric spatial modelling with Dirichlet process mixing. *Journal of the American Statistical Association* **100**, 1021–1035.

Griffin, J.E., and Steel, M.F.J. (2006). Order based dependent Dirichlet processes. *Journal of the American Statistical Association* **101**, 179–194.

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, **7**, 223–242.

Hanson, T. (2006). Inference for mixtures of finite Polya tree models. *Journal of the American Statistical Association*, **101**, 1548–1565.

Hanson, T., Branscum, A., and Gardner, I. (2008). Multivariate mixtures of Polya trees for modelling ROC data. *Statistical Modelling*, **8**, 81–96.

Jara, A. and Hanson, T. (2011). A class of mixtures of dependent tailfree processes. *Biometrika*, **98**, 553–566.

Jo, S., Lee, J., Müller, P., Quintana, F., and Trippa, L. (2017). Dependent species sampling model for spatial density estimation. *Bayesian Analysis*, **12**, 379–406.

MacEachern, S.N. (2001). Decision theoretic aspects of dependent nonparametric processes. In *Bayesian Methods with Applications to Science, Policy and Official Statistics* (Edited by E. George), 551–560. Eurostat.

Newton, M.A., and Zhang, Y. (1999). A recursive algorithm for nonparametric analysis with missing data. *Biometrika*, **86**, 15–26.

Petrone, S., Guindani, M., and Gelfand, A.E. (2009). Hybrid Dirichlet mixture models for functional data. *Journal of the Royal Statistical Society, Ser. B* **71**, 755–782.

Reich, B., and Fuentes, M. (2007). A multivariate semiparametric bayesian spatial modeling framework for hurricane surface wind fields. *Annals of Applied Statistics* **1**, 249–264.

Rodríguez, A., Dunson, D., and Gelfand, A. (2010). Latent stick-breaking processes. *Journal of the American Statistical Association*, **105**, 647–659.

Schörgendorfer, A., and Branscum, A.J. (2013). Regression analysis using dependent Polya trees. *Statistics in Medicine*, **32**, 4679–4695.

Tansey, W., Athey, A., Reinhart, A., and Scott, J.G. (2017). Multiscale spatial density smoothing: an application to large-scale radiological survey and anomaly detection. *Journal of the American Statistical Association*, **112**, 1047–1063.

Zhao, L., and Hanson, T. (2011). Spatially dependent Polya tree modeling for survival data. *Biometrics*, **67**, 391–403.

Zhou, H., Hanson, T., and Knapp, R. (2015). Marginal Bayesian nonparametric model for time to disease arrival of threatened amphibian populations. *Biometrics*, **71**, 1101–1110.

Zhou, H., and Hanson, T. (2018). ***spBayesSurv**: Bayesian Modeling and Analysis of Spatially Correlated Survival Data*. R package version 1.1.3 or higher, URL `http://CRAN.R-project.org/package=spBayesSurv`.

Zhou, H., Hanson, T., and Zhang, J. (2017). Generalized accelerated failure time spatial frailty model for arbitrarily censored data. *Lifetime Data Analysis*, **23**, 495–515.