A unified framework for fitting Bayesian semiparametric models to arbitrarily censored survival data, including spatially-referenced data

Haiming Zhou
Division of Statistics, Northern Illinois University
and
Timothy Hanson
Department of Statistics, University of South Carolina

July 3, 2017

(*To appear in Journal of the American Statistical Association*)

Abstract

A comprehensive, unified approach to modeling arbitrarily censored spatial survival data is presented for the three most commonly-used semiparametric models: proportional hazards, proportional odds, and accelerated failure time. Unlike many other approaches, all manner of censored survival times are simultaneously accommodated including uncensored, interval censored, current-status, left and right censored, and mixtures of these. Left-truncated data are also accommodated leading to models for time-dependent covariates. Both georeferenced (location exactly observed) and areally observed (location known up to a geographic unit such as a county) spatial locations are handled; formal variable selection makes model selection especially easy. Model fit is assessed with conditional Cox-Snell residual plots, and model choice is carried out via LPML and DIC. Baseline survival is modeled with a novel transformed Bernstein polynomial prior. All models are fit via a new function which calls efficient compiled C++ in the R package `spBayesSurv`. The methodology is broadly illustrated with simulations and real data applications. An important finding is that proportional odds and accelerated failure time models often fit significantly better than the commonly-used proportional hazards model. Supplementary materials are available online.

Keywords: Bernstein polynomial; Interval censored data; Spatial frailty models; Variable selection

1 Introduction

Spatial location often plays a key role in prediction, serving as a proxy for unmeasured regional characteristics such as socioeconomic status, access and quality of healthcare, pollution, etc. Spatial models use location both as a means for blocking, leading to more precisely estimated non-spatial risk factors, but also as a focal point of inference in its own right, for example to delineate regional "hot spots" or outbreaks that merit closer attention or increased resources. Literature on the spatial analysis of time-to-event data related to human health has flourished over the last decade, including data on leukemia survival (Henderson et al., 2002), infant/childhood mortality (Banerjee et al., 2003; Kneib, 2006), coronary artery bypass grafting (Hennerfeind et al., 2006), asthma (Li and Ryan, 2002; Li and Lin, 2006), breast cancer (Zhao and Hanson, 2011; Hanson et al., 2012; Zhou et al., 2015), mortality due to air pollution (Jerrett et al., 2013), colorectal cancer survival (Liu et al., 2014), smoking cessation (Pan et al., 2014), HIV/AIDS patients (Martins et al., 2016), time to tooth loss (Schnell et al., 2015), and many others. Spatial survival models have also been used in other important areas such as the study of political event processes (Darmofal, 2009), agricultural mildew outbreaks (Ojiambo and Kang, 2013), forest fires (Morin, 2014), pine trees (Li et al., 2015a), health and pharmaceutical firms (Arbia et al., 2017), and emergency service response times (Taylor, 2017) to name a few.

All twenty papers referenced above use Cox's (Cox, 1972) proportional hazards (PH) model for inference; competing models are not considered. There are a few papers using models alternative to PH in a spatial context, e.g. Diva et al. (2008), Zhao et al. (2009), Wang et al. (2012), and Li et al. (2015b) but they tend to be limited in scope. For example, all four of these latter papers only consider areal data, do not employ variable selection or allow time-dependent covariates, and are developed for right censored data only. In general the literature is fragmented in that a method for variable selection in the PH model for non-spatial right censored data will comprise one paper; another paper may extend the PH model for fitting current status data to general interval censored data, but does not include variable

selection, spatial frailties, etc. In this paper, a broadly inclusive, comprehensive treatment of three competing, highly interpretable semiparametric survival models, PH, proportional odds (PO), and accelerated failure time (AFT), is developed and illustrated on several real and simulated data sets. This work represents the culmination of a great deal of effort tying together many disparate ideas and methodologies in the literature as well as developing an efficient approach to mixed interval censored, exactly observed, and truncated data that can be widely applied to different semiparametric models.

Model formulation consists of a parametric portion giving relative risk (PH), odds (PO), or acceleration factors (AFT) coupled with georeferenced or areal spatial frailties, and a nonparametric baseline survival function modeled as a transformed Bernstein polynomial (TBP) (Chen et al., 2014) centered at a standard parametric family: one of log-logistic, log-normal, or Weibull. Centering the baseline allows prior mass to be roughly guided by the parametric family. The resulting distribution behaves similarly to a B-spline but where knot locations are guided by the centering family, blending the merits of both parametric and nonparametric approaches. Unlike mixtures of Dirichlet processes (Antoniak, 1974), gamma processes (Kalbfleisch, 1978), and mixtures of Polya trees (Lavine, 1992), the TBP is smooth and has a finite number of parameters, yielding an explicit likelihood that allows for many immediate generalizations and efficient block-adaptive Markov chain Monte Carlo (MCMC) inference. With judicious choice of blocking and careful use of existing parametric survival model fits, the block-adaptive approach cultivated here is robust, fully automated (e.g. no "tuning" is required), is relatively fast, and has been applied to data sets of up to a million observations. Furthermore, all of the machinery developed is available in a powerful, freely-available R function `survregbayes` calling compiled C++ in the `spBayesSurv` package (Zhou and Hanson, 2017) for R. Note that although the methodology is developed for both areal and georeferenced spatial time-to-event data, non-spatial data are also accommodated. All manner of exactly observed, right censored, interval censored, and left-truncated data are accommodated, as well as mixtures of these. Left-truncation allows for the inclusion of

time-dependent covariates and spike-and-slab variable selection is also implemented. Finally, a separate function that computes a variation on the Cox-Snell residual plot allows for gross assessment of model fit. The ready availability of software to easily fit the models developed herein allows researchers to empirically compare various competing semiparametric models on their own survival data, as well as allowing comparison to other survival models in the literature.

Section 2 describes the models including the TBP, georeferenced and areal frailties, MCMC, diagnostics and model selection criteria. Three illustrative data analyses comprise Section 3. Section 4 offers simulations illustrating the quality of estimation as well as comparing to the R packages `ICBayes` (Pan et al., 2015), `bayesSurv` (Komárek and Lesaffre, 2008) and `R2BayesX` (Umlauf et al., 2015; Belitz et al., 2015). The paper is concluded in Section 5. Tests for parametric baseline, stochastic search variable selection, left-truncation, time-dependent covariates, partially linear predictors, more simulations, as well as many other details are further discussed in the online supplementary material accompanying this paper.

2 The Models

Subjects are observed at m distinct spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_m$. For areally-observed outcomes, e.g. county-level, there is typically replication at each location; for georeferenced data there may or may not be replication. Let t_{ij} be the (possibly censored) survival time for subject j at location $\mathbf{s}_i$ and $\mathbf{x}_{ij}$ be the corresponding p-dimensional vector of covariates, $i = 1, \ldots, m, j = 1, \ldots, n_i$; let $n = \sum_{i=1}^{m} n_i$ be the total number of subjects. Assume the survival time t_{ij} lies in the interval (a_{ij}, b_{ij}), $0 \leq a_{ij} \leq b_{ij} \leq \infty$. Here left censored data are of the form $(0, b_{ij})$, right censored (a_{ij}, ∞), interval censored (a_{ij}, b_{ij}) and uncensored values simply have $a_{ij} = b_{ij}$, i.e. define $(x, x) = \{x\}$. The event and censoring times are assumed to be independent given the observed covariates. Note that both current status data and

case 2 interval censored data (Sun, 2006), arise as particular special cases.

Spatial dependence often arises among survival outcomes due to region-specific similarities in ecological and/or social environments that are typically not measurable or omitted due to confidentiality concerns. To incorporate such spatial dependence, a traditional way is to introduce random effects (frailties) $v_1, \ldots, v_m$ into the linear predictor of survival models. In this paper we consider three commonly-used semiparametric models: AFT, PH, and PO. Given spatially-varying frailties $v_1, \ldots, v_m$, regression effects $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, and baseline survival $S_0(\cdot)$ with density $f_0(\cdot)$ corresponding to $\mathbf{x}_{ij} = \mathbf{0}$ and $v_i = 0$, the AFT model has survival and density functions

$$S_{\mathbf{x}_{ij}}(t) = S_0(e^{\mathbf{x}_{ij}'\boldsymbol{\beta}+v_i}t),\ f_{\mathbf{x}_{ij}}(t) = e^{\mathbf{x}_{ij}'\boldsymbol{\beta}+v_i}f_0(e^{\mathbf{x}_{ij}'\boldsymbol{\beta}+v_i}t), \tag{2.1}$$

while the PH model has survival and density functions

$$S_{\mathbf{x}_{ij}}(t) = S_0(t)^{e^{\mathbf{x}_{ij}'\boldsymbol{\beta}+v_i}},\ f_{\mathbf{x}_{ij}}(t) = e^{\mathbf{x}_{ij}'\boldsymbol{\beta}+v_i}S_0(t)^{e^{\mathbf{x}_{ij}'\boldsymbol{\beta}+v_i}-1}f_0(t), \tag{2.2}$$

and the PO model has survival and density functions

$$S_{\mathbf{x}_{ij}}(t) = \frac{e^{-\mathbf{x}_{ij}'\boldsymbol{\beta}-v_i}S_0(t)}{1+(e^{-\mathbf{x}_{ij}'\boldsymbol{\beta}-v_i}-1)S_0(t)},\ f_{\mathbf{x}_{ij}}(t) = \frac{e^{-\mathbf{x}_{ij}'\boldsymbol{\beta}-v_i}f_0(t)}{[1+(e^{-\mathbf{x}_{ij}'\boldsymbol{\beta}-v_i}-1)S_0(t)]^2}. \tag{2.3}$$

In semiparametric survival analysis, a wide variety of Bayesian nonparametric priors can be used to model $S_0(\cdot)$; see Müller et al. (2015) and Zhou and Hanson (2015) for reviews. In this paper we consider the models (2.1), (2.2) and (2.3) with a TBP prior on $S_0(\cdot)$.

2.1 Transformed Bernstein Polynomial Prior

For a given positive integer $J \geq 1$, define the Bernstein polynomial of degree $J-1$ as a particular B-spline over $(0,1)$ given by

$$d(x|J, \mathbf{w}_J) = \sum_{j=1}^{J} w_j \delta_{j,J}(x) \equiv \sum_{j=1}^{J} w_j \frac{\Gamma(J+1)}{\Gamma(j)\Gamma(J-j+1)} x^{j-1}(1-x)^{J-j}, \tag{2.4}$$

where $\mathbf{w}_J = (w_1, \ldots, w_J)'$ is a vector of positive weights satisfying $\sum_{j=1}^{J} \mathbf{w}_j = 1$ and $\delta_{j,J}(x)$ denotes a beta density with parameters $(j, J-j+1)$. Smooth densities with support $(0,1)$ can be well approximated by a Bernstein polynomial (Ghosal, 2001): if $f(x)$ is any continuously differentiable density with support $(0,1)$ and bounded second derivative, $\mathbf{w}_J$ can be chosen such that

$$\sup_{0<x<1} |f(x) - d(x|J, \mathbf{w}_J)| = O(J^{-1}).$$

Integrating (2.4) gives the corresponding cumulative distribution function (cdf)

$$D(x|J, \mathbf{w}_J) = \sum_{j=1}^{J} w_j \Delta_{j,J}(x), \tag{2.5}$$

where $\Delta_{j,J}(x)$ is the cdf associated with $\delta_{j,J}(x)$. One can calculate the summands in (2.5) recursively as

$$\Delta_{j+1,J}(x) = \Delta_{j,J}(x) - \frac{\Gamma(J+1)}{\Gamma(j+1)\Gamma(J-j+1)} x^{j}(1-x)^{J-j}.$$

By assigning a joint prior distribution to $(J, \mathbf{w}_J)$, the random $d(x|J, \mathbf{w}_J)$ in (2.4) is said to have the Bernstein polynomial (BP) prior. Petrone (1999) showed that if the prior on $(J, \mathbf{w}_J)$ has full support, the BP prior has positive support on all continuous density functions on $(0,1)$. However, for practical reasons, the degree J is often truncated to a large value, say $\mathcal{K}$, so that the prior has support $\mathcal{B}_{\mathcal{K}} = \{d(x|J, \mathbf{w}_J) : J \leq \mathcal{K}\}$. Under mild conditions, Petrone and Wasserman (2002) showed that, for a fixed $\mathcal{K}$, the posterior density almost

surely converges (as $n \to \infty$) to a density that minimizes the Kullback-Leibler divergence of the true density against $d(x) \in \mathcal{B}_{\mathcal{K}}$. BP priors of lesser degree have the same support as $\mathcal{B}_{\mathcal{K}}$ because any Bernstein polynomial can be written in terms of Bernstein polynomials of higher degree through the relationship

$$\delta_{j,J-1}(x) = \frac{J-j}{J}\delta_{j,J}(x) + \frac{j}{J}\delta_{j+1,J}(x).$$

It follows that $d(x|J-1, \mathbf{w}_{J-1})$ can be written as $d(x|J, \mathbf{w}_J^*)$ with a suitable choice of $\mathbf{w}_J^*$, so every $d(x|J, \mathbf{w}_J)$ with $J \leq \mathcal{K}$ belongs to $\{d(x|\mathcal{K}, \mathbf{w}_{\mathcal{K}})\}$. Therefore, following Chen et al. (2014) we fix J throughout, with $J = 15$ being the software's default. This is roughly equivalent to having 15 knots in a B-spline representation of the transformed baseline survival function, described next.

Let $\{S_{\boldsymbol{\theta}}(\cdot) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ denote a parametric family of survival functions with support on positive reals $\mathbb{R}^+$. The log-logistic family $S_{\boldsymbol{\theta}}(t) = \{1 + (e^{\theta_1}t)^{\exp(\theta_2)}\}^{-1}$ is considered in the examples in Section 3, where $\boldsymbol{\theta} = (\theta_1, \theta_2)'$; Weibull and log-normal families are also included in the R package. Note that $S_{\boldsymbol{\theta}}(t)$ always lies in the interval $(0, 1)$ for $0 < t < \infty$, so a natural prior on $S_0(\cdot)$, termed the TBP prior, is simply

$$S_0(t) = D(S_{\boldsymbol{\theta}}(t)|J, \mathbf{w}_J) \text{ with density } f_0(t) = d(S_{\boldsymbol{\theta}}(t)|J, \mathbf{w}_J) f_{\boldsymbol{\theta}}(t), \tag{2.6}$$

where $f_{\boldsymbol{\theta}}$ is the density associated with $S_{\boldsymbol{\theta}}(\cdot)$. Clearly, the random distribution $S_0(\cdot)$ is centered at $S_{\boldsymbol{\theta}}(\cdot)$, that is, $E[S_0(t)] = S_{\boldsymbol{\theta}}(t)$ and $E[f_0(t)] = f_{\boldsymbol{\theta}}(t)$. The weight parameters $\mathbf{w}_J$ adjust the shape of the baseline survival $S_0(\cdot)$ relative to the centering distribution $S_{\boldsymbol{\theta}}(\cdot)$. This adaptability makes the TBP prior attractive in its flexibility, but also anchors the random $S_0(\cdot)$ firmly about $S_{\boldsymbol{\theta}}(\cdot)$: $w_j = 1/J$ for $j = 1, \ldots, J$ implies $S_0(t) = S_{\boldsymbol{\theta}}(t)$ for $t \geq 0$. Moreover, unlike the mixture of Polya trees (Lavine, 1992) or mixture of Dirichlet process (Antoniak, 1974) priors, the TBP prior selects smooth densities, leading to efficient posterior sampling. Initially we had implemented the models using mixtures of Polya trees (and this

function is still available in the `spBayesSurv` R package with limited functionality); these models suffered poor mixing, especially for the AFT model. The TBP provided essentially the same posterior inference or better (as measured by LPML) with vastly improved MCMC mixing; see supplemental Appendix J.4.

Regarding the prior for $\mathbf{w}_J$, we consider a Dirichlet distribution, $\mathbf{w}_J|J \sim \text{Dirichlet}(\alpha, \ldots, \alpha)$, where $\alpha > 0$ acts like the precision in a Dirichlet process (Ferguson, 1973), controlling how stochastically "pliable" $S_0(\cdot)$ is relative to $S_{\boldsymbol{\theta}}(\cdot)$. Large values of α indicate a strong belief that $S_0(\cdot)$ is close to $S_{\boldsymbol{\theta}}(\cdot)$: as $\alpha \to \infty$, $S_0(t) \to S_{\boldsymbol{\theta}}(t)$ with probability 1. Smaller values of α allow more pronounced deviations of $S_0(\cdot)$ from $S_{\boldsymbol{\theta}}(\cdot)$. A gamma prior on α is considered: $\alpha \sim \Gamma(a_\alpha, b_\alpha)$ where $E(\alpha) = a_\alpha/b_\alpha$.

2.2 Spatial Frailty Modeling

Spatial frailty models are usually grouped into two general settings according to their underlying data structure (Banerjee et al., 2014): *georeferenced* data, where $\mathbf{s}_i$ varies continuously throughout a fixed study region $\mathcal{S}$ (e.g., $\mathbf{s}_i$ is recorded as longitude and latitude); and *areal* data, where $\mathcal{S}$ is partitioned into a finite number of areal units with well-defined boundaries (e.g., $\mathbf{s}_i$ represents a county).

2.2.1 Areal Data Modeling

We consider an intrinsic conditionally autoregressive (ICAR) prior (Besag, 1974) on $\mathbf{v} = (v_1, \ldots, v_m)'$. Let e_{ij} be 1 if regions i and j are neighbors (which can be defined in various ways) and 0 otherwise; set $e_{ii} = 0$. Then the $m \times m$ matrix $\mathbf{E} = [e_{ij}]$ is called the adjacency matrix for the m regions. The ICAR prior on $\mathbf{v}$ is defined through the set of the conditional distributions

$$v_i|\{v_j\}_{j\neq i} \sim N\left(\sum_{j=1}^{m} e_{ij}v_j/e_{i+},\ \tau^2/e_{i+}\right),\ i = 1, \ldots, m, \tag{2.7}$$

where $e_{i+} = \sum_{j=1}^{m} e_{ij}$, τ is a scale parameter, and τ^2/e_{i+} is the conditional variance. The induced prior on $\mathbf{v}$ under ICAR is improper; the constraint $\sum_{j=1}^{m} v_j = 0$ is used for identifiability (Banerjee et al., 2014). Note that we assume that every region has at least one neighbor, so the proportionality constant for the improper density of $\mathbf{v}$ is $(\tau^{-2})^{(m-1)/2}$ (Lavine and Hodges, 2012).

A referee has asked about the inclusion of the proper CAR prior, which is not currently supported by the software accompanying this article. The ICAR has been generally favored over its proper counterpart, the CAR prior, for several reasons. The CAR prior includes a parameter κ which shrinks the frailties toward zero: $v_i|\kappa, \{v_j\}_{j\neq i} \sim N\left(\kappa \sum_{j=1}^{m} e_{ij}v_j/e_{i+}, \tau^2/e_{i+}\right)$. As $\kappa \to 1^-$ the improper ICAR results. Paciorek (2009) suggests that the shrinkage from a proper CAR is "generally unappealing" in the spatial setting; similarly Banerjee et al. (2014) question the sensibility of smoothing a spatial effect toward a proportion $\kappa < 1$ of the mean of its neighbors. Paciorek (2009) further notes that posterior estimates of κ are often close to one, essentially yielding the ICAR. This is likely due to the rather modest amount of correlation the proper CAR provides unless $\kappa \approx 1$; see Assunção and Krainski (2009) and Banerjee et al. (2014).

When spatial smoothing is not of interest, we consider independent and identically distributed (IID) Gaussian frailties, $v_1, \ldots, v_m \overset{iid}{\sim} N(0, \tau^2)$. This is a special case of the proper CAR prior where $\kappa = 0$, so the software allows for either no correlation or the maximal limiting correlation of the proper CAR prior.

2.2.2 Georeferenced Data Modeling

For georeferenced data, it is commonly assumed that $v_i = v(\mathbf{s}_i)$ arises from a Gaussian random field (GRF) $\{v(\mathbf{s}), \mathbf{s} \in \mathcal{S}\}$ such that $\mathbf{v} = (v_1, \ldots, v_m)$ follows a multivariate Gaussian distribution as $\mathbf{v} \sim N_m(\mathbf{0}, \tau^2\mathbf{R})$, where τ^2 measures the amount of spatial variation across locations and the (i, j) element of $\mathbf{R}$ is modeled as $\mathbf{R}[i, j] = \rho(\mathbf{s}_i, \mathbf{s}_j)$, where $\rho(\cdot, \cdot)$ is a correlation function controlling the spatial dependence of $v(\mathbf{s})$. In this paper, we consider

the powered exponential correlation function $\rho(\mathbf{s}, \mathbf{s}') = \rho(\mathbf{s}, \mathbf{s}'; \phi) = \exp\{-(\phi\|\mathbf{s} - \mathbf{s}'\|)^\nu\}$, where $\phi > 0$ is a range parameter controlling the spatial decay over distance, $\nu \in (0, 2]$ is a shape parameter, and $\|\mathbf{s} - \mathbf{s}'\|$ is the distance (e.g., Euclidean, great-circle) between $\mathbf{s}$ and $\mathbf{s}'$. Note that $\phi \to \infty$ gives the IID Gaussian prior described in Section 2.2.1 as a special case. Similar to ICAR, the conditional prior is given by

$$v_i|\{v_j\}_{j\neq i} \sim N\left(-\sum_{\{j:j\neq i\}} p_{ij}v_j/p_{ii},\ \tau^2/p_{ii}\right),\ i = 1, \ldots, m, \tag{2.8}$$

where p_{ij} is the (i, j) element of $\mathbf{R}^{-1}$.

As m increases evaluating $\mathbf{R}^{-1}$ from $\mathbf{R}$ becomes computationally impractical. Various approaches have been developed to overcome this computational issue such as process convolutions (Higdon, 2002), fixed rank kriging (Cressie and Johannesson, 2008), predictive process models (Banerjee et al., 2008), sparse approximations (Kaufman et al., 2008), and the full-scale approximation (Sang and Huang, 2012). We consider the full-scale approximation (FSA) as an option when fitting models in the accompanying R software due to its capability of capturing both large- and small-scale spatial dependence; see supplementary Appendix B for a brief introduction.

The FSA was arrived at after a number of lengthy, unsuccessful attempts at using other approximations. BayesX (Belitz et al., 2015) uses what have been termed "Matérn splines," first introduced in an applied context by Kammann and Wand (2003). Several authors have used this approach including Kneib (2006), Hennerfeind et al. (2006), and Kneib and Fahrmeir (2007). This approximation was termed a "predictive process" and given a more formal treatment by Banerjee et al. (2008). In our experience, the predictive process tends to give biased regression effects and prediction when the rank (i.e. the number of knots) was chosen too low; the problem worsened with no replication and/or when spatial correlation was high. The FSA fixes the predictive process by adding tapering to the "residual" process. We have been able to successfully analyze data with several thousand georeferenced spatial

locations via MCMC using the FSA option with little to no artificial bias in parameter estimation and consistently good predictive ability.

Choices alternative to the powered exponential such as the spherical and Matérn correlation functions may be of interest, however the current package only supports the powered exponential correlation with pre-specified ν. This choice has proven to be adequate across many simulated data sets and is partially borne from our experience with very weakly identified parameters in "too flexible" correlation structures. Initially we used the Matérn but moved to the powered exponential when Matérn parameters would invariably have very poor MCMC mixing and biased estimates in all but very large samples. Note that the exponential correlation function, also a special case of Matérn, is given by $\nu = 1$ yielding continuous but not differentiable sample paths. Gaussian correlation, a limiting case of Matérn, is given by $\nu = 2$, providing (infinitely) differentiable sample paths.

BayesX and R-INLA (Martins et al., 2013) both include the Matérn correlation function $\rho(\mathbf{s}, \mathbf{s}'; \phi) = [2^{\nu-1}\Gamma(\nu)]^{-1}(\phi\|\mathbf{s} - \mathbf{s}'\|)^{\nu}\mathcal{K}_{\nu}(\phi\|\mathbf{s} - \mathbf{s}'\|)$ where $\mathcal{K}_{\nu}$ is the modified Bessel function of order ν. Noting weakly identified Matérn parameters, BayesX requires the user to fix one of $\nu \in \{0.5, 1.5, 2.5, 3.5\}$ ($\nu = 0.5$ gives exponential correlation; $\nu \to \infty$ gives Gaussian) and fixes ϕ^{-1} at $\hat{\phi}^{-1} = \max_{i,j} ||\mathbf{s}_i - \mathbf{s}_j||/c$ with c chosen so the correlation between the two farthest locations is small, e.g. 0.001. R-INLA requires the user to fix one of $\nu \in \{1.5, 2.5, 3.5\}$ but allows the range parameter ϕ to be estimated from the data. So the approach of BayesX fixes both parameters of the Matérn correlation and R-INLA allows the range ϕ to be random (as do we).

2.3 Likelihood Construction and MCMC

Let $\mathcal{D} = \{(a_{ij}, b_{ij}, \mathbf{x}_{ij}, \mathbf{s}_i); i = 1, \ldots, m, j = 1, \ldots, n_i\}$ be the observed data. Assume $t_{ij} \sim S_{\mathbf{x}_{ij}}(t)$ following one of (2.1), (2.2), or (2.3) with the TBP prior on $S_0(t)$ defined in (2.6),

and $\mathbf{v}$ following (2.7) or (2.8). The likelihood for $(\mathbf{w}_J, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v})$ is

$$L(\mathbf{w}_J, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}) = \prod_{i=1}^{m} \prod_{j=1}^{n_i} \left[S_{\mathbf{x}_{ij}}(a_{ij}) - S_{\mathbf{x}_{ij}}(b_{ij}) \right]^{I\{a_{ij} < b_{ij}\}} f_{\mathbf{x}_{ij}}(a_{ij})^{I\{a_{ij} = b_{ij}\}}. \tag{2.9}$$

MCMC is carried out through an empirical Bayes approach coupled with adaptive Metropolis samplers (Haario et al., 2001). The posterior density is

$$p(\mathbf{w}_J, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}, \alpha, \tau^2, \phi | \mathcal{D}) \propto L(\mathbf{w}_J, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}) p(\mathbf{w}_J | \alpha) p(\alpha) p(\boldsymbol{\theta}) p(\boldsymbol{\beta}) p(\mathbf{v} | \tau^2) p(\tau^2) p(\phi),$$

where each $p(\cdot)$ represents a prior density, and $p(\phi)$ is only included for georeferenced data. Assume $\boldsymbol{\theta} \sim N_2(\boldsymbol{\theta}_0, \mathbf{V}_0)$, $\boldsymbol{\beta} \sim N_p(\boldsymbol{\beta}_0, \mathbf{W}_0)$, $\alpha \sim \Gamma(a_\alpha, b_\alpha)$, $\tau^{-2} \sim \Gamma(a_\tau, b_\tau)$ and $\phi \sim \Gamma(a_\phi, b_\phi)$.

Recall that $w_j = 1/J$ implies the underlying parametric model with $S_0(t) = S_{\boldsymbol{\theta}}(t)$. Thus, the parametric model provides good starting values for the TBP survival model. Let $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$ denote the parametric estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, and let $\hat{\mathbf{V}}$ and $\hat{\mathbf{W}}$ denote their estimated covariance matrices, respectively. Set $\mathbf{z}_{J-1} = (z_1, \ldots, z_{J-1})'$ with $z_j = \log(w_j) - \log(w_J)$. The $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, $\mathbf{z}_{J-1}$, α and ϕ are all updated using adaptive Metropolis samplers, where the initial proposal variance is $\hat{\mathbf{W}}$ for $\boldsymbol{\beta}$, $\hat{\mathbf{V}}$ for $\boldsymbol{\theta}$, $0.16\mathbf{I}_{J-1}$ for $\mathbf{z}_{J-1}$ and 0.16 for α and ϕ. Each frailty term v_i is updated via Metropolis-Hastings, with proposal variance as the conditional prior variance of $v_i | \{v_j\}_{j \neq i}$; τ^{-2} is updated via a Gibbs step from its full conditional. A complete description and derivation of the updating steps are in supplementary Appendix A. To determine the running length of an MCMC run, one may first run a short chain without thinning, then use R packages such as `coda` (Plummer et al., 2006) and `mcmcse` (Flegal et al., 2016) for convergence diagnosis and effective sample size calculations.

Regarding the default choice for the hyperparameters, when variable selection is not implemented (see Appendix E in the online supplementary material) we set $\boldsymbol{\beta}_0 = \mathbf{0}$, $\mathbf{W}_0 = 10^{10}\mathbf{I}_p$, $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}$, $\mathbf{V}_0 = 10\hat{\mathbf{V}}$, $a_\alpha = b_\alpha = 1$, and $a_\tau = b_\tau = .001$. Note here we assume a somewhat informative prior on $\boldsymbol{\theta}$ to obviate confounding between $\boldsymbol{\theta}$ and $\mathbf{w}_J$. For georeferenced data, we set $a_\phi = 2$ and $b_\phi = (a_\phi - 1)/\phi_0$ so that the prior of ϕ has mode at ϕ_0, where ϕ_0

satisfies $\rho(\mathbf{s}', \mathbf{s}''; \phi_0) = 0.001$ with $\|\mathbf{s}' - \mathbf{s}''\| = \max_{i,j} \|\mathbf{s}_i - \mathbf{s}_j\|$. Note that Kneib and Fahrmeir (2007) simply fix ϕ at ϕ_0, while we allow ϕ to be random around ϕ_0.

2.4 Model Diagnostics and Comparison

For model diagnostics, a general residual defined in Cox and Snell (1968) has been widely used in a variety of regression settings. Define $r(t_{ij}) = -\log S_{\mathbf{x}_{ij}}(t_{ij})$ for $i = 1, \ldots, m$, $j = 1, \ldots, n_i$, then $r(t_{ij})$, given $S_{\mathbf{x}_{ij}}(\cdot)$, has a standard exponential distribution. Therefore, if the model is "correct," and under arbitrary censoring, the pairs $\{r(a_{ij}), r(b_{ij})\}$ are approximately a random arbitrarily censored sample from an Exp(1) distribution, and the estimated (Turnbull, 1974) cumulative hazard plot should be approximately straight with slope 1. Uncertainty in the plot is assessed through several cumulative hazards based on a random sample from $[\mathbf{w}_J, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}|\mathcal{D}]$. This is in contrast to typical Cox-Snell plots which only use point estimates.

Several researchers have pointed out that Cox-Snell residuals are conservative in that they may be straight even under quite large departures from the model (Baltazar-Aban and Pena, 1995; O'Quigley and Xu, 2005). In this case, model criteria will be more informative. We consider two popular model choice criteria: the deviance information criterion (DIC) (Spiegelhalter et al., 2002) and the log pseudo marginal likelihood (LPML) (Geisser and Eddy, 1979), where DIC (smaller is better) places emphasis on the relative quality of model fitting and LPML (larger is better) focuses on the predictive performance. Both criteria are readily computed from the MCMC output; see supplemental Appendix C for more details.

3 Real Data Applications

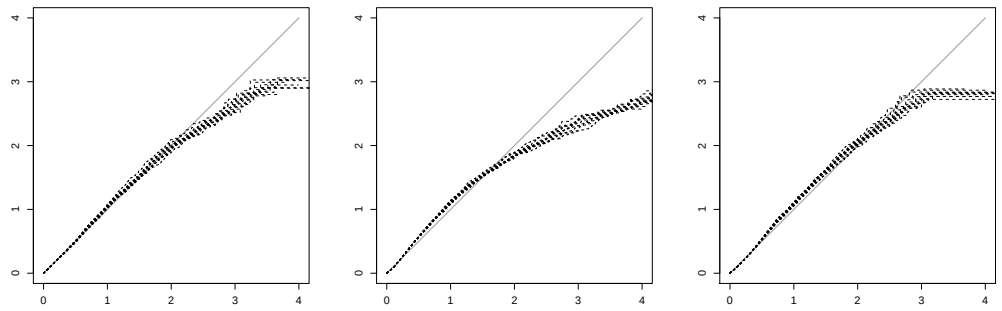
3.1 Loblolly Pine Survival Data

Loblolly pine is the most commercially important timber species in the Southeastern United States; estimating loblolly survival is crucial to forestry research. The dataset used in this

section consists of 45,525 loblolly pine trees at $m = 168$ distinct sites, which were established in 1980-1981 and monitored annually until 2001-2002. During the 21-year follow-up, 5,379 trees died; the rest survived until the last follow-up and are treated as right censored. It is of interest to investigate the association between loblolly pine survival and several important risk factors after adjusting for spatial dependence among different sites. The risk factors considered include two time-independent variables, treatment and physiographic region, and three time-dependent variables—diameter at breast height, tree height and crown class—which were repeatedly measured every 3 years. Supplementary Table S6 presents some baseline characteristics for the trees.

Li et al. (2015a) fitted a semiparametric PH model with several georeferenced spatial frailty specifications. However, they showed that the PH assumption does not hold very well for treatment and physiographic region. To investigate whether the AFT or PO provides better fit, we fit each of the AFT, PH and PO models with GRF frailties to the data using the same covariates as those in Li et al. (2015a); the log-logistic centering distribution was used throughout. Here $\nu = 1$ was fixed giving the exponential correlation function. To better investigate the spatial frailty effect, we also fit each model with IID Gaussian frailties and non-frailty models. For each MCMC run we retained $2,000$ scans thinned from $50,000$ after a burn-in period of $10,000$ iterations. Figure 1 reports the Cox-Snell residual plots under the three GRF frailty models, where we see that the PH model severely deviates from the 45 degree line, and the AFT fits the data much better than the PH and PO. Table 1 compares all fitted models using the LPML and DIC criteria, where we can see that the AFT always outperforms the PH and PO regardless of the frailty assumptions. Under all three models, incorporating IID frailties significantly improves the goodness of fit over the non-frailty model. Some goodness of fit improvement for the GRF over the IID occurs under the AFT, but this does not happen under the PH and PO, indicating that adding spatially structured frailties can deteriorate the model if the overarching model assumption is violated.

Table 2 gives covariate effects under the AFT models. Coefficient estimates under the



(a) (b) (c)

Figure 1: Loblolly pine data. Cox-Snell residual plots under AFT (panel a), PH (panel b) and PO (panel c) with GRF frailties.

Table 1: Loblolly pine data. Model comparison.

		AFT	PH	PO
GRF frailty	LPML	-23,812	-23,991	-23,882
	DIC	47,611	47,971	47,767
IID frailty	LPML	-23,832	-23,966	-23,865
	DIC	47,648	47,897	47,731
Non-frailty	LPML	-25,447	-25,508	-25,549
	DIC	50,893	51,015	51,099

model with frailties (IID or GRF) have changed significantly from those under the non-frailty model. For example, the effect of total tree height on tree survival is reversed when the model is fit with frailties. This indicates that shorter trees are associated with longer survival rates when averaged over spatial location; however taller trees have better survival rates than shorter adjusting for location. Thus a type of "Simpson's paradox" occurs with space being confounded with tree height in some fashion.

We next interpret the results under the AFT model for time-dependent covariates (Prentice and Kalbfleisch, 1979) with GRF frailties, as it outperforms all other models. Table 2 shows all covariates are significant risk factors for loblolly pine survival. For example, the mean or median survival time will increase by a factor $e^{0.126} = 1.134$ for every 1 cm increase in diameter at breast height, holding other covariates and the frailty constant. Hanson et al. (2009) note that this interpretation holds for mean or median *residual life* as well, which is of greater interest in medical contexts when choosing a course of treatment. Significant

Table 2: Loblolly pine data. Posterior means (95% credible intervals) of fixed effects $\boldsymbol{\beta}$ from fitting the AFT model with different frailty settings.

	Non-frailty	IID frailty	GRF frailty
DBH	-0.233(-0.255,-0.212)	-0.127(-0.144,-0.111)	-0.126(-0.142,-0.110)
TH	0.027(0.024,0.030)	-0.012(-0.015,-0.010)	-0.011(-0.014,-0.009)
treat2	-0.521(-0.583,-0.466)	-0.381(-0.423,-0.339)	-0.388(-0.431,-0.349)
treat3	-0.719(-0.798,-0.641)	-0.529(-0.589,-0.473)	-0.544(-0.601,-0.495)
PhyReg2	-0.302(-0.367,-0.241)	-0.362(-0.514,-0.198)	-0.390(-0.594,-0.201)
PhyReg3	-0.007(-0.110,0.099)	-0.254(-0.530,0.050)	-0.260(-0.511,0.014)
C2	0.097(0.028,0.169)	0.031(-0.023,0.078)	0.044(-0.002,0.097)
C3	0.835(0.747,0.923)	0.401(0.331,0.465)	0.430(0.375,0.491)
C4	1.919(1.799,2.029)	1.040(0.951,1.128)	1.101(1.018,1.194)
treat2:PhyReg2	0.146(0.041,0.244)	0.118(0.053,0.190)	0.105(0.046,0.168)
treat3:PhyReg2	0.372(0.246,0.503)	0.255(0.169,0.349)	0.246(0.162,0.332)
treat2:PhyReg3	-0.404(-0.625,-0.197)	-0.220(-0.374,-0.072)	-0.216(-0.368,-0.064)
treat3:PhyReg3	0.109(-0.129,0.325)	0.110(-0.067,0.267)	0.125(-0.037,0.286)
τ^2		0.318(0.248,0.402)	0.350(0.270,0.457)
ϕ			0.274(0.165,0.434)

interaction effects are present between treatment and physiographic region; Figure 2 presents the survival curves for the treatment effect under different physiographic regions. The thinning treatment has the largest effect on survival rates in coastal regions, while for Piedmont regions, heavy thinning is required to improve survival rates. The posterior means of spatial frailties for each location are also mapped in Figure 2, where we do not see a clear spatial pattern, implying that the spatial dependence may not be very strong for these data. To further confirm this, the posterior mean of the spatial range parameter ϕ is 0.274 km. Note that 99% of the pairwise distances among the 168 locations are greater than 9.93 km, that is, 99% of the pairwise correlations are lower than $1 - e^{-0.274 \times 9.93} \approx 0.066$.

3.2 The Signal Tandmobiel Study

Oral health data collected from the Signal Tandmobiel study are next considered, available in the R package `bayesSurv`; a description of the data can be found in Komárek and Lesaffre (2009). It is of interest to investigate the impact of gender (1 = girl, 0 = boy), dmf (1, if

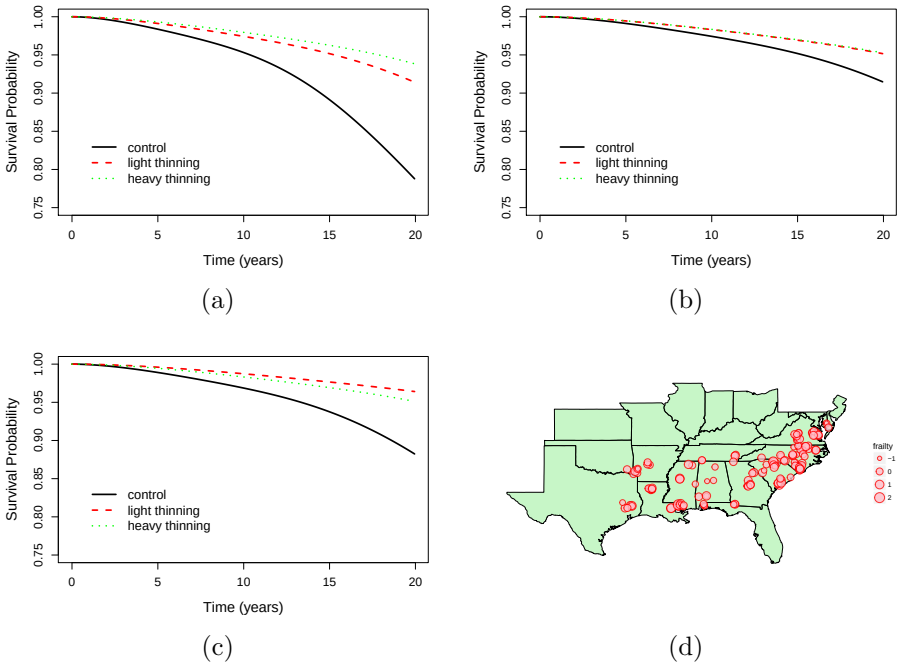


Figure 2: Loblolly pine data. Survival curves for treatment effect under coastal (panel a), Piedmont (panel b), and other (panel c) regions. The posterior means of spatial frailties for each location are mapped in panel d.

the predecessor of the permanent first premolar was decayed, missing due to caries or filled, 0, if the predecessor was sound), and tooth location (1 = mandibular, 0 = maxillary) on the emergence time of each of the four permanent first premolar (teeth 14, 24, 34, 44). The data set consists of $4,430$ children with four tooth emergence times recorded for each child, yielding a sample size of $n = 17,594$. The emergence times are interval censored due to annual examinations. Komárek and Lesaffre (2009) fit an IID frailty AFT model for taking into account the dependency between emergence times within each child. Although their models allow for a nonparametric frailty density, their Figure 6 shows a remarkably Gaussian-shaped estimate. Each of the AFT, PH and PO models with IID Gaussian frailties were fit to the data using the same covariates as the Model S in (Komárek and Lesaffre, 2009, Table 4), giving LPML values -15125, -15124, and -15503, respectively. Thus, the PH and AFT models perform similarly and both outperform the PO in terms of predictive performance; however, the AFT model fits these data better than the PH according to DIC. Cox-Snell

Table 3: The Signal Tandmobiel study. Posterior means (95% credible intervals) of fixed effects from fitting the AFT, PH and PO models with IID frailties. The LPML and DIC are also shown for each model.

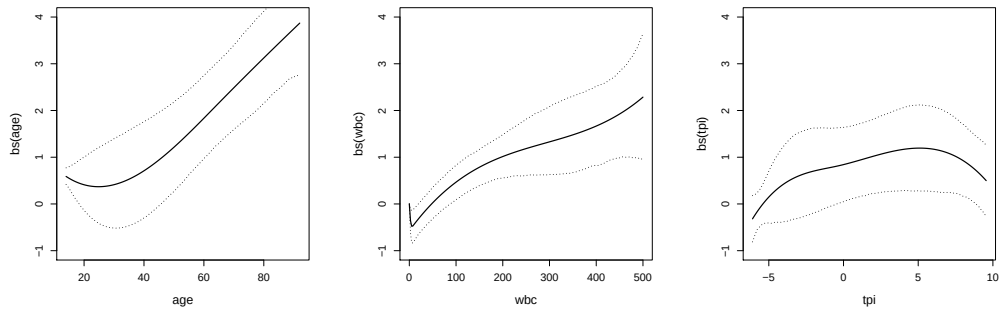
	AFT (LPML: -15125) (DIC: 29067)	PH (LPML: -15124) (DIC: 30045)	PO (LPML: -15503) (DIC: 30596)
gender	0.045(0.038,0.053)	1.034(0.884,1.191)	1.433(1.206,1.667)
dmf	0.031(0.026,0.036)	0.799(0.690,0.908)	1.424(1.264,1.573)
tooth	0.020(0.017,0.022)	0.408(0.341,0.472)	0.594(0.496,0.694)
dmf:tooth	-0.018(-0.022,-0.014)	-0.478(-0.578,-0.378)	-0.770(-0.921,-0.625)
gender:dmf	-0.009(-0.015,-0.003)	-0.258(-0.401,-0.126)	-0.470(-0.674,-0.260)
τ^2	0.010(0.009, 0.010)	4.619(4.326, 4.918)	9.261(8.679, 9.885)

residual plots (not shown) also slightly favor AFT. Table 3 reports the fixed effects and frailty variances under each model, from which we see that the effect of dmf is different for boys and girls, and it is also different for mandibular and maxillary teeth.

3.3 Leukemia Data

Finally, a dataset on the survival of acute myeloid leukemia in $n = 1,043$ patients is considered, available in the package `spBayesSurv`. It is of interest to investigate possible spatial variation in survival after accounting for known subject-specific prognostic factors, which include age, sex, white blood cell count (wbc) at diagnosis, and the Townsend score (tpi) for which higher values indicates less affluent areas. There are $m = 24$ administrative districts and each one forms a spatial cluster (Henderson et al., 2002, Figure 1). Henderson et al. (2002) fitted a multivariate gamma frailty PH model with linear predictors. Here we fit each of the PH, AFT and PO models with ICAR frailties to see whether the AFT or PO model provides better fit. To allow for non-linear effects of continuous predictors, we consider partially linear age, wbc and tpi as described in supplementary Appendix G.

The LPML values for the PH, AFT and PO models are -5946, -5945, and -5919, respectively. The PO model significantly outperforms others from a predictive point of view with a pseudo Bayes factor on the order of 10^{10} relative to PH and AFT. Figure 3 presents the



(a) (b) (c)

Figure 3: Leukemia data. Posterior mean estimates (solid lines) for the effects of age (panel a), wbc (panel b) and tpi (panel c), together with 95% credible intervals (dotted lines).

estimated effects of age, wbc and tpi under the PO model. The Bayes Factors for testing the linearity of age, wbc and tpi are 0.13, 0.04 and 0.01, respectively; non-linear effects are not needed.

4 Simulations

Extensive simulations were carried out to evaluate the proposed MCMC algorithms (implemented in `spBayesSurv`) under the three survival models with arbitrarily censored spatial data. The proposed methodology is then compared to monotone splines (Lin et al., 2015) as implemented in `ICBayes` for PH and PO, Bayesian G-splines (Komarek, 2006) implemented in `bayesSurv` for AFT, and geoadditive PH modeling (Hennerfeind et al., 2006) implemented in `R2BayesX`. Additional simulations for georeferenced data and variable selection are available in supplementary Appendixes J.2 and J.3.

4.1 Simulation I: Areal Data

For each of AFT, PH and PO, data were generated from (2.1), (2.2) and (2.3), respectively, where covariates $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2})'$ were sampled $x_{ij1} \overset{iid}{\sim}$ Bernoulli(0.5) independent of $x_{ij2} \overset{iid}{\sim} N(0,1)$ for $i = 1 \ldots, 37$ and $j = 1, \ldots, 20$ ($n = 740$), regression effects set to $\boldsymbol{\beta} = (\beta_1, \beta_2)' = (1,1)'$, the baseline survival distribution $S_0(t) = 1 - 0.5[\Phi(2(\log t + 1)) + \Phi(2(\log t - 1))]$

was bimodal, and $\mathbf{v}$ followed the ICAR model (2.7) with $\tau^2 = 1$ and $\mathbf{E}$ equaling the Nigeria adjacency matrix used in the childhood mortality data analysis of Kneib (2006). Half the data were right censored (including uncensored survival times) and half interval censored. The times at which survival was right censored were independently simulated from a Uniform$(2, 6)$ distribution. For interval censoring, each subject was assumed to have N observation times, $O_1, O_2, \ldots, O_N$, where $(N-1) \sim$ Poisson(2) and $(O_k - O_{k-1})|N \overset{iid}{\sim}$ Exp(1) with $O_0 = 0$, $k = 1, \ldots, N$. The censoring interval endpoints are the two adjacent observation times among $\{0, O_1, \ldots, O_N, \infty\}$ that include the true survival time. The final data yield around 20% right censored, 40% uncensored, 25% left censored and 15% interval censored. For each model, 500 Monte Carlo (MC) replicate data sets were generated. Models were fit using the default priors introduced in Section 2.3. For each MCMC run, $5,000$ scans were thinned from $50,000$ after a burn-in period of $10,000$ iterations; convergence diagnostics deemed this more than adequate.

Table 4 summarizes the results for regression parameters $\boldsymbol{\beta}$ and the ICAR variance τ^2, including the averaged bias (BIAS) and posterior standard deviation (PSD) of each point estimate (posterior mean for $\boldsymbol{\beta}$ and median for τ^2), the standard deviation (across 500 MC replicates) of the point estimate (SD-Est), the coverage probability (CP) of the 95% credible interval, and average effective sample size (ESS) out of 5,000 (Sargent et al., 2000) for each point estimate. The results show that the point estimates of $\boldsymbol{\beta}$ and τ^2 are unbiased under all three models, SD-Est values are close to the corresponding PSDs, the CP values are close to the nominal 95% level, and ESS values are promising. Supplementary Figure S1 presents the average across the 500 MC replicates of the fitted baseline survival functions revealing that the proposed model is capable to capture complex baseline survival curves very well.

4.2 Simulation II: Comparison

In this section we compare our R function `survregbayes` in the package `spBayesSurv` with R functions `ICBayes` in the package `ICBayes`, `bayessurvreg2` in the package `bayesSurv` and

Table 4: Simulation I: average bias (BIAS) and posterior standard deviation (PSD) of each point estimate, standard deviation (across 500 MC replicates) of the point estimate (SD-Est), coverage probability (CP) for the 95% credible interval, and effective sample size (ESS) out of 5,000 with thinning=10 for each point estimate.

Model	Parameter	BIAS	PSD	SD-Est	CP	ESS
AFT	$\beta_1 = 1$	0.000	0.067	0.062	0.960	3194
	$\beta_2 = 1$	0.002	0.036	0.034	0.960	2992
	$\tau^2 = 1$	0.015	0.311	0.298	0.942	4561
PH	$\beta_1 = 1$	-0.021	0.100	0.099	0.936	3024
	$\beta_2 = 1$	-0.014	0.061	0.060	0.944	2095
	$\tau^2 = 1$	-0.039	0.352	0.319	0.950	3478
PO	$\beta_1 = 1$	0.012	0.151	0.152	0.940	3672
	$\beta_2 = 1$	0.008	0.083	0.086	0.944	2822
	$\tau^2 = 1$	-0.008	0.465	0.448	0.924	2451

`bayesx` in the package `R2BayesX` in terms of mixing and computing speed. Our `survregbayes` function can fit spatial or non-spatial PH, AFT and PO models (either no, IID, ICAR or GRF frailties) for all common types of survival (uncensored, interval censored, current status, right censored, etc.) and/or left-truncated data. In contrast, `ICBayes` fits only non-frailty PH and PO models to interval censored data (uncensored data is not supported), `bayessurvreg2` fits the AFT model (either no or IID frailties) to general interval censored data, and `bayesx` fits the PH model to right-censored data using MCMC. Data are generated from three cases: (C1) the non-frailty PH model with interval censoring, (C2) the non-frailty AFT model with arbitrary censoring, and (C3) the non-frailty PH model with right censoring, where $\boldsymbol{\beta}$ and $S_0(\cdot)$ are the same as those used in **Simulation I**. Under each setting we generate 500 MC replicates, each with sample size $n = 500$. For each MCMC run, we retain $10,000$ scans without thinning after a burn-in period of $10,000$ iterations.

Table 5 reports the comparison results; the ESS from `survregbayes` range from 3 to 20 times as large as those using `ICBayes` and `bayessurvreg2`, indicating that the proposed MCMC algorithms are more efficient in terms of mixing. In addition, `survregbayes` is about 5 times faster than `ICBayes`, although it is much slower than `bayessurvreg2` and `bayesx`. Note that `bayessurvreg2` is almost 9 times faster than our `survregbayes`, but its ESS for

Table 5: Simulation II: average bias (BIAS) and posterior standard deviation (PSD) of each point estimate, standard deviation (across 500 MC replicates) of the point estimate (SD-Est), coverage probability (CP) for the 95% credible interval, average effective sample size (ESS) out of 10,000 without thinning for each point estimate, and average computing time in seconds.

Case	R function	Time	Parameter	BIAS	PSD	SD-Est	CP	ESS
C1	`survregbayes`	63	$\beta_1 = 1$	-0.018	0.134	0.134	0.940	1139
			$\beta_2 = 1$	-0.015	0.086	0.087	0.940	934
	`ICBayes`	310	$\beta_1 = 1$	-0.036	0.133	0.132	0.938	346
			$\beta_2 = 1$	-0.019	0.084	0.085	0.938	292
C2	`survregbayes`	54	$\beta_1 = 1$	0.000	0.075	0.074	0.950	1207
			$\beta_2 = 1$	0.002	0.041	0.040	0.952	1186
	`bayessurvreg2`	6	$\beta_1 = 1$	-0.002	0.072	0.075	0.936	64
			$\beta_2 = 1$	-0.001	0.039	0.040	0.934	109
C3	`survregbayes`	85	$\beta_1 = 1$	-0.009	0.103	0.061	0.954	1028
			$\beta_2 = 1$	-0.010	0.107	0.061	0.950	673
	`bayesx`	44	$\beta_1 = 1$	0.017	0.103	0.062	0.940	989
			$\beta_2 = 1$	0.014	0.111	0.064	0.952	2817

β_1 is nearly 19 times smaller than ours. That is, `bayessurvreg2` needs to take 19 times more iterates than ours to obtain the same ESS, thus tempering superior speed with much poorer MCMC mixing. In comparison with `bayesx` for right censored data, our `survregbayes` performs slightly worse in terms of both speed and mixing. This is not surprising, because our MCMC is designed for all AFT, PH and PO with all manner of censoring schemes while the MCMC in `bayesx` is tailored to PH with right censored data.

`ICBayes` stems from a burgeoning literature on fitting PO and PH models to interval censored data using a monotone spline, essentially an integrated B-spline, as the baseline cumulative hazard function. In particular Cai et al. (2011) consider a PH model for current status data; Lin and Wang (2011) consider the PO model with current status data; case 2 interval censoring is considered for PO by Wang and Lin (2011). Pan et al. (2014) consider the PH model for general interval censored data with spatial CAR frailties and Lin et al. (2015) consider the same PH model with general interval censored data without frailties.

Wang et al. (2016) take the E-M approach to the model of Lin et al. (2015). Among these papers Lin et al. (2015) state concerning the `ICBayes` function that "*...our developed Gibbs sampler is easy to execute with only four steps and efficient because it does not require imputing any unobserved failure times or contain any complicated Metropolis-Hastings steps.*" They further say that their approach is efficient in the sense that their approach is easy to program. However, although the Lin et al. (2015) approach does not require imputing failure times, it does require imputing as many latent Poisson variates as there are spline basis functions per observation, i.e. much more latent data than simply imputing the survival times. Furthermore, they update one regression coefficient β_j at a time for $j = 1, \ldots, p$ via the ARS method (Gilks and Wild, 1992), and it is well-known that any component-at-a-time sampler will suffer from poor mixing if there is strong correlation in the posterior. The simplest way to improve mixing is through the consideration correlated proposals such as the adaptive version developed here or the iteratively weighted least squares proposal used in Hennerfeind et al. (2006). This simulation shows that adaptive block updates can dramatically outperform a Gibbs sampler with latent data, both in terms of speed and mixing. Furthermore, the monotone spline approach models the baseline cumulative hazard over the range of the observed data only; however, the AFT model maps observed data onto "baseline support" through a scale factor. The TBP prior used here includes a scale and so is applicable to all of PH, PO, and AFT, as well as accelerated hazards (which also needs a scale term), additive hazards, proportional mean residual life, and others.

5 Discussion

The recent surge in literature analyzing spatially correlated time-to-event data focusing on PH highlights the need for flexible, robust methodology and software to enable the ready use of other competing yet easily-interpretable survival models, spatial or not. This article provides a unified framework for considering competing models to PH, i.e. AFT and PO,

including variable selection, areal or georeferenced frailties, and additive structure. The freely available `survregbayes` function allows for easy fitting of these competing models. An important aspect associated with the Bayesian nonparametric TBP – or more accurately richly parametric – formulation of the AFT, PH, and PO models presented here is that the assumption of the same flexible prior on $S_0(\cdot)$ places these models common ground. Differences in fit and/or predictive performance can therefore be attributed to the semiparametric models only, rather than to additional possible differences in quite different nonparametric priors (e.g. gamma process vs. beta process vs. Dirichlet process mixture) or estimation methods (e.g. NPMLE vs. partial likelihood vs. sieves).

In two of the three data analyses in Section 3 a model alternative to PH was favored by the predictive LPML measure; for the dental data in Section 3.2, the AFT and PH predictively perform about equally well and outperform PO. However, PH can be certainly be the best model choice from a predictive point of view; this is, of course, data-dependent. In additional analyses not presented here, the proposed AFT, PH and PO models were applied to spatial smoking cessation data set from the Lung Health Study carried out by Murray et al. (1998) and analyzed by Pan et al. (2014). The LPML values were -211, -206, and -206, respectively, indicating that PH and PO models predict equally well and outperform AFT; the pseudo Bayes factor comparing PH to AFT (or PO to AFT) is approximately $e^{-206-(-211)} \approx 150$. We also analyzed the current status lung cancer data set available in the R package `ICBayes`, analyzed by Sun (2006). The LPML values under the AFT, PH and PO are all essentially -83, respectively, indicating that there is essentially no difference on the predictive performance among the three models. This makes sense as the only covariate (treatment) was not significant: $\boldsymbol{\beta} = \mathbf{0}$ implies PH, PO, and AFT reduce to $S_0(\cdot)$.

Supplementary Materials

Appendices: Appendix A: MCMC sampling algorithms; Appendix B: brief introduction to the full-scale approximation; Appendix C: definitions of the DIC and LPML criteria; Appendix D: testing for parametric $S_0(\cdot)$; Appendix E: stochastic search variable selection; Appendix F: left-truncation and time-dependent covariates; Appendix G: partially linear predictors; Appendix H: introduction to the R function `survregbayes`; Appendix I: additional results for real data applications; Appendix J: additional simulation results. (pdf file)

Data and Code: Data and R code for the three analyses in Section 3. (zip file)

References

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 2(6):1152–1174.

Arbia, G., Espa, G., Giuliani, D., and Micciolo (2017). A spatial analysis of health and pharmaceutical firm survival. *Journal of Applied Statistics*, 44(9):1560–1575.

Assunção, R. and Krainski, E. (2009). Neighborhood dependence in Bayesian spatial models. *Biometrical Journal*, 51(5):851–869.

Baltazar-Aban, I. and Pena, E. A. (1995). Properties of hazard-based residuals and implications in model diagnostics. *Journal of the American Statistical Association*, 90(429):185–197.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data, Second Edition.* Chapman and Hall/CRC Press.

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process

models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848.

Banerjee, S., Wall, M. M., and Carlin, B. P. (2003). Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics*, 4(1):123–142.

Belitz, C., Brezger, A., Klein, N., Kneib, T., Lang, S., and Umlauf, N. (2015). *BayesX - Software for Bayesian Inference in Structured Additive Regression Models*. Version 3.0. Available from http://www.bayesx.org.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B*, 36(2):192–236.

Cai, B., Lin, X., and Wang, L. (2011). Bayesian proportional hazards model for current status data with monotone splines. *Computational Statistics & Data Analysis*, 55(9):2644–2651.

Chen, Y., Hanson, T., and Zhang, J. (2014). Accelerated hazards model based on parametric families generalized with Bernstein polynomials. *Biometrics*, 70(1):192–201.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, 30(2):248–275.

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):209–226.

Darmofal, D. (2009). Bayesian spatial survival models for political event processes. *American Journal of Political Science*, 53(1):241–257.

Diva, U., Dey, D. K., and Banerjee, S. (2008). Parametric models for spatially correlated survival data for individuals with multiple cancers. *Statistics in Medicine*, 27(12):2127–2144.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230.

Flegal, J. M., Hughes, J., and Vats, D. (2016). *mcmcse: Monte Carlo Standard Errors for MCMC*. Riverside, CA and Minneapolis, MN. R package version 1.2-1.

Geisser, S. and Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74(365):153–160.

Ghosal, S. (2001). Convergence rates for density estimation with Bernstein polynomials. *The Annals of Statistics*, 29(5):1264–1280.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41(2):337–348.

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242.

Hanson, T., Johnson, W., and Laud, P. (2009). Semiparametric inference for survival models with step process covariates. *Canadian Journal of Statistics*, 37(1):60–79.

Hanson, T. E., Jara, A., Zhao, L., et al. (2012). A Bayesian semiparametric temporally-stratified proportional hazards model with spatial frailties. *Bayesian Analysis*, 7(1):147–188.

Henderson, R., Shimakura, S., and Gorst, D. (2002). Modeling spatial variation in leukemia survival data. *Journal of the American Statistical Association*, 97(460):965–972.

Hennerfeind, A., Brezger, A., and Fahrmeir, L. (2006). Geoadditive survival models. *Journal of the American Statistical Association*, 101(475):1065–1075.

Higdon, D. (2002). Space and space-time modeling using process convolutions. In *Quantitative methods for current environmental issues*, pages 37–56. Springer.

Jerrett, M., Burnett, R. T., Beckerman, B. S., Turner, M. C., Krewski, D., Thurston, G., Martin, R. V., van Donkelaar, A., Hughes, E., Shi, Y., Gapstur, S. M., Thun, M. J., and Pope, C. A. (2013). Spatial analysis of air pollution and mortality in California. *American Journal of Respiratory and Critical Care Medicine*, 188(5):593–599.

Kalbfleisch, J. (1978). Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society, Series B*, 40(2):214–221.

Kammann, E. E. and Wand, M. P. (2003). Geoadditive models. *Applied Statistics*, 52(1):1–18.

Kaufman, C. G., Schervish, M. J., and Nychka, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103(484):1545–1555.

Kneib, T. (2006). Mixed model-based inference in geoadditive hazard regression for interval-censored survival times. *Computational Statistics & Data Analysis*, 51(2):777–792.

Kneib, T. and Fahrmeir, L. (2007). A mixed model approach for geoadditive hazard regression. *Scandinavian Journal of Statistics*, 34(1):207–228.

Komarek, A. (2006). *Accelerated failure time models for multivariate interval-censored data with flexible distributional assumptions*. PhD thesis, PhD thesis, Katholieke Universiteit Leuven, Faculteit Wetenschappen.

Komárek, A. and Lesaffre, E. (2008). Bayesian accelerated failure time model with multivariate doubly-interval-censored data and flexible distributional assumptions. *Journal of the American Statistical Association*, 103(482):523–533.

Komárek, A. and Lesaffre, E. (2009). The regression analysis of correlated interval-censored data illustration using accelerated failure time models with flexible distributional assumptions. *Statistical Modelling*, 9(4):299–319.

Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modelling. *The Annals of Statistics*, 20(3):1222–1235.

Lavine, M. L. and Hodges, J. S. (2012). On rigorous specification of ICAR models. *The American Statistician*, 66(1):42–49.

Li, J., Hong, Y., Thapa, R., and Burkhart, H. E. (2015a). Survival analysis of loblolly pine trees with spatially correlated random effects. *Journal of the American Statistical Association*, 110(510):486–502.

Li, L., Hanson, T., and Zhang, J. (2015b). Spatial extended hazard model with application to prostate cancer survival. *Biometrics*, 71(2):313–322.

Li, Y. and Lin, X. (2006). Semiparametric normal transformation models for spatially correlated survival data. *Journal of the American Statistical Association*, 101(474):591–603.

Li, Y. and Ryan, L. (2002). Modeling spatial survival data using semiparametric frailty models. *Biometrics*, 58(2):287–297.

Lin, X., Cai, B., Wang, L., and Zhang, Z. (2015). A Bayesian proportional hazards model for general interval-censored data. *Lifetime Data Analysis*, 21(3):470–490.

Lin, X. and Wang, L. (2011). Bayesian proportional odds models for analyzing current status data: Univariate, clustered, and multivariate. *Communications in Statistics-Simulation and Computation*, 40(8):1171–1181.

Liu, Y., Sun, D., and He, C. Z. (2014). A hierarchical conditional autoregressive model for colorectal cancer survival data. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(1):37–44.

Martins, R., Silva, G. L., and Andreozzi, V. (2016). Bayesian joint modeling of longitudinal and spatial survival AIDS data. *Statistics in Medicine*, 35(19):3368–3384.

Martins, T. G., Simpson, D., Lindgren, F., and Rue, H. (2013). Bayesian computing with INLA: New features. *Computational Statistics & Data Analysis*, 67:68 – 83.

Morin, A. A. (2014). A spatial analysis of forest fire survival and a marked cluster process for simulating fire load. Master's thesis, The University of Western Ontario, London, Ontario, Canada.

Müller, P., Quintana, F., Jara, A., and Hanson, T. (2015). *Bayesian Nonparametric Data Analysis*. Springer-Verlag: New York.

Murray, R. P., Anthonisen, N. R., Connett, J. E., Wise, R. A., Lindgren, P. G., Greene, P. G., Nides, M. A., Group, L. H. S. R., et al. (1998). Effects of multiple attempts to quit smoking and relapses to smoking on pulmonary function. *Journal of clinical epidemiology*, 51(12):1317–1326.

Ojiambo, P. and Kang, E. (2013). Modeling spatial frailties in survival analysis of cucurbit downy mildew epidemics. *Phytopathology*, 103(3):216–227.

O'Quigley, J. and Xu, R. (2005). Goodness of fit in survival analysis. In *Encyclopedia of Biostatistics*. John Wiley & Sons, Ltd.

Paciorek, C. (2009). Technical Vignette 5: Understanding intrinsic Gaussian Markov random field spatial models, including intrinsic conditional autoregressive models. Technical report, Department of Statistics, University of California, Berkeley, and Department of Biostatistics, Harvard School of Public Health. http://www.stat.berkeley.edu/~paciorek/research/techVignettes/techVignette5.pdf.

Pan, C., Cai, B., Wang, L., and Lin, X. (2014). Bayesian semiparametric model for spa-

tially correlated interval-censored survival data. *Computational Statistics & Data Analysis*, 74:198–209.

Pan, C., Cai, B., Wang, L., and Lin, X. (2015). *ICBayes: Bayesian Semiparametric Models for Interval-Censored Data*. R package version 1.0.

Petrone, S. (1999). Random Bernstein polynomials. *Scandinavian Journal of Statistics*, 26(3):373–393.

Petrone, S. and Wasserman, L. (2002). Consistency of Bernstein polynomial posteriors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(1):79–100.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *Journal of the American Statistical Association*, 6(1):7–11.

Prentice, R. L. and Kalbfleisch, J. D. (1979). Hazard rate models with covariates. *Biometrics*, 35(1):25–39.

Sang, H. and Huang, J. Z. (2012). A full scale approximation of covariance functions for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(1):111–132.

Sargent, D. J., Hodges, J. S., and Carlin, B. P. (2000). Structured Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics*, 9(2):217–234.

Schnell, P., Bandyopadhyay, D., Reich, B. J., and Nunn, M. (2015). A marginal cure rate proportional hazards model for spatial survival data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(4):673–691.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, 64(4):583–639.

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*. Springer-Verlag New York.

Taylor, B. M. (2017). Spatial modelling of emergency service response times. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(2):433–453.

Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association*, 69(345):169–173.

Umlauf, N., Adler, D., Kneib, T., Lang, S., and Zeileis, A. (2015). Structured additive regression models: An R interface to BayesX. *Journal of Statistical Software*, 63(21):1–46.

Wang, L. and Lin, X. (2011). A Bayesian approach for analyzing case 2 interval-censored data under the semiparametric proportional odds model. *Statistics & Probability Letters*, 81(7):876–883.

Wang, L., McMahan, C. S., Hudgens, M. G., and Qureshi, Z. P. (2016). A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. *Biometrics*, 72(1):222–231.

Wang, S., Zhang, J., and Lawson, A. B. (2012). A Bayesian normal mixture accelerated failure time spatial model and its application to prostate cancer. *Statistical Methods in Medical Research*, http://dx.doi.org/10.1177/0962280212466189.

Zhao, L. and Hanson, T. E. (2011). Spatially dependent Polya tree modeling for survival data. *Biometrics*, 67(2):391–403.

Zhao, L., Hanson, T. E., and Carlin, B. P. (2009). Mixtures of Polya trees for flexible spatial frailty survival modelling. *Biometrika*, 96(2):263–276.

Zhou, H. and Hanson, T. (2015). Bayesian spatial survival models. In *Nonparametric Bayesian Inference in Biostatistics*, pages 215–246. Springer.

Zhou, H. and Hanson, T. (2017). *spBayesSurv: Bayesian Modeling and Analysis of Spatially Correlated Survival Data*. R package version 1.1.1.

Zhou, H., Hanson, T., Jara, A., and Zhang, J. (2015). Modeling county level breast cancer survival data using a covariate-adjusted frailty proportional hazards model. *The Annals of Applied Statistics*, 9(1):43–68.