Bayesian Spatial Survival Models

Haiming Zhou and Timothy Hanson

1 Introduction

This chapter reviews several semiparametric Bayesian survival models, and summarizes some recent proposals to allow for spatial and covariate-adjusted dependence among the survival times. Two generalizations of the accelerated failure time model that allow crossing cumulative hazards for different covariate combinations, and hence crossing survival curves, are also discussed.

Four prior specifications in broad use are first reviewed in Section 2. A catalogue of Bayesian survival models is presented in Section 3. Section 4 discusses the incorporation of dependence among survival times across the models in Section 3, focusing mostly on spatial dependence followed by several real-data illustrations in Section 5. The chapter concludes with a short discussion in Section 6. Please note at the outset that, although a review is attempted, the cited papers and approaches are biased toward what the authors are aware of and have found useful.

2 A selection of nonparametric priors

A common starting point in the specification of a regression model for time-to-event data is the definition of a baseline survival function, S_0, that is modified (either directly or indirectly) by subject-specific covariates $\mathbf{x}$. Let T_0 be a random survival time from the baseline group (with all covariates equal to zero). The baseline survival function is defined by $S_0(t) = P(T_0 > t) = \exp\{-H_0(t)\}$ where $H_0(t)$ is the

Haiming Zhou
Northern Illinois University Division of Statistics, DeKalb IL 60115 e-mail: zhouh@stat.sc.edu

Timothy Hanson
University of South Carolina Department of Statistics, Columbia SC 29208

baseline cumulative hazard. For continuous outcomes, the baseline density and hazard functions are $f_0(t) = -\frac{d}{dt}S_0(t)$ and $h_0(t) = f_0(t)/S_0(t) = \frac{d}{dt}H_0(t)$, respectively. The cumulative distribution, survival, density and hazard functions for a member of the population with covariates $\mathbf{x}$ will be denoted by $F_{\mathbf{x}}(t)$, $S_{\mathbf{x}}(t)$, $f_{\mathbf{x}}(t)$, and $h_{\mathbf{x}}(t)$, respectively.

A wide variety of priors have been used in Bayesian survival analysis over the last 40 years. We focus on four of these: the gamma process, B-splines, Dirichlet process mixtures, and mixtures of Polya trees. Additional reviews can be found in Sinha and Dey (1997), Ibrahim et al. (2001), Müller and Quintana (2004), Hanson et al. (2005), Nieto-Barajas (2013), and Müller et al. (2015).

2.1 Gamma process

Kalbfleisch (1978) proposed the gamma process (GP) to model the cumulative hazard function H_0 in the context of the proportional hazards (PH) model (Cox, 1972). Let $H_\theta(t)$ be an increasing, left-continuous function on $[0,\infty)$ indexed by θ, where $H_\theta(0) = 0$; typically H_θ is parametric. Let $H_0(\cdot)$ be a stochastic process such that (i) $H_0(0) = 0$, (ii) $H_0(\cdot)$ has independent increments in disjoint intervals, and (iii) $H_0(t_2) - H_0(t_1) \sim \Gamma\{\alpha(H_\theta(t_2) - H_\theta(t_1)), \alpha\}$ for $t_2 > t_1$, where $\Gamma(\alpha,\beta)$ implies mean α/β. Then $\{H_0(t) : t \geq 0\}$ is said to be a GP with parameter (α, H_θ) and denoted $H_0 \sim GP(\alpha, H_\theta)$.

Note that $\mathrm{E}\{H_0(t)\} = H_\theta(t)$ so that H_0 is centered at H_θ. Also, $\mathrm{Var}\{H_0(t)\} = H_\theta(t)/\alpha$ so that, similar to the Dirichlet process and Polya trees described below, the precision parameter α controls how "close" H_0 is to H_θ and provides a prior measure of how certain one is that H_0 is near H_θ. Ferguson (1973) recast the Dirichlet process (DP) as a scaled GP.

The posterior of the GP is characterized by Kalbfleisch (1978); his results for the PH model simplify when no covariates are specified. With probability one, the GP is a monotone nondecreasing step function, implying that the corresponding survival function S_0 is a nonincreasing step function. Similar to the DP, matters are complicated by the presence of ties in the data with positive probability. When present in the observed data, such ties make the resulting computations awkward. Clayton (1991) described a Gibbs sampler for obtaining inferences in the PH model with a GP baseline.

Burridge (1981) and Ibrahim et al. (2001) suggest that the model as proposed by Kalbfleisch (1978) and extended by Clayton (1991) is best suited to grouped survival data. Walker and Mallick (1997) considered an approximation to the GP for continuous data. Define a partition of $(0,\infty)$ by $\{(a_{j-1}, a_j]\}_{j=1}^J \cup (a_J, \infty)$ where $0 = a_0 < a_1 < a_2 < \cdots < a_{J+1} = \infty$. Here, a_J is taken to be equal to be largest event time recorded. If $H_0 \sim GP(\alpha, H_\theta)$ then by definition $h_{0j} = H_0(a_j) - H_0(a_{j-1}) \overset{ind.}{\sim} \Gamma\{\alpha(H_\theta(a_j) - H_\theta(a_{j-1})), \alpha\}$. Walker and Mallick (1997) make this assumption for the given partition and further assume that $h_0(t)$ is constant and equal to h_{0j} for $t \in (a_{j-1}, a_j]$, $j = 1,\ldots,J$, yielding a particular piecewise exponential model. So the

piecewise exponential model, which has a long and fruitful history in both Bayesian and frequentist survival analysis, can be viewed as an approximation to the GP when gamma increments are used.

2.2 B-splines and Bernstein polynomials

A flexible and popular basis expansion approach to modeling functions over a finite interval $[a,b]$ is based on B-splines (de Boor, 2001). A B-spline is a piecewise-differentiable polynomial of a given degree d; $d = 2$ and $d = 3$ give quadratic and cubic B-splines, respectively. The B-spline is defined over the union of intervals with endpoints termed knots. The overall polynomial is continuous ($d \geq 1$) or differentiable ($d \geq 2$) over the range of the knots. Knots can be equispaced yielding a cardinal B-spline or else irregularly-spaced. Computation is especially easy for equispaced knots and so we focus on that here; generalizations can be found in Kneib (2006). The B-spline includes polynomials of the same or lower degree as special cases; e.g. a quadratic B-spline includes all constant, linear, and parabolic functions over $[a,b]$.

For degree $d = 2$, the quadratic B-spline "mother" basis function is defined on $[0,3]$

$$\varphi(x) = \left\{ \begin{array}{ll} 0.5x^2 & 0 \leq x \leq 1 \\ 0.75 - (x-1.5)^2 & 1 \leq x \leq 2 \\ 0.5(3-x)^2 & 2 \leq x \leq 3 \\ 0 & \text{otherwise} \end{array} \right\}.$$

Say the number of basis functions is J. The B-spline basis functions are shifted, rescaled versions of φ. Let $x_1, \ldots, x_n$ be event times of interest and $x_{(1)}, \ldots, x_{(n)}$ their order statistics. The j-th basis function is $B_j(x) = \varphi\left(\frac{x - x_{(1)}}{\Delta} + 3 - j\right)$, where $\Delta = \frac{x_{(n)} - x_{(1)}}{J-2}$. A B-spline is typically used with a rather large number of basis functions J, e.g. 20–40. The B-spline model for an unknown function is

$$g(x) = \sum_{j=1}^{J} \theta_j B_j(x). \tag{1}$$

A global level of smoothness can be incorporated into a B-spline model by encouraging neighboring coefficients to be similar; the more regular the coefficients are, the less wiggly g is. The hazard can be modeled directly as $h_0(t) = g(t)$ with the constraint $\theta_j \geq 0$ (Wang and Dunson, 2011; Pan et al., 2014; Lin et al., 2015; Li et al., 2015b); typically $\theta_1, \ldots, \theta_J$ have exponential or gamma priors. Komárek and Lesaffre (2008) consider a limiting case of the B-spline order as a model for densities and model the $\theta_j \geq 0$ via a generalized logit transformation so that $\sum_{j=1}^{J} \theta_j = 1$.

Alternatively, to avoid the positivity constraints on θ_i, one can model $h_0(t) = \exp\{g(t)\}$ (Hennerfeind et al., 2006; Kneib and Fahrmeir, 2007) with $\theta_j \in \mathbb{R}$. Classical spline estimation on $\{(x_i, y_i)\}_{i=1}^{n}$ proceeds by minimizing $\sum_{i=1}^{n} (y_i - g(x_i))^2$

subject to the "wiggliness" penalty $\int_a^b |g''(x)|^2 dx \leq c$ for some $c > 0$. This is equivalent to maximizing a penalized log-likelihood. Borrowing from Eilers and Marx (1996), Lang and Brezger (2004) recast and developed this idea into a Bayesian framework. Let $\mathbf{D}_2 \in \mathbb{R}^{(J-2)\times J}$ and $\mathbf{D}_1 \in \mathbb{R}^{(J-1)\times J}$ be defined as

$$\mathbf{D}_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix} \text{ and } \mathbf{D}_1 = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{bmatrix}.$$

For equispaced, quadratic (and cubic) B-splines the penalty can be written as $\int_a^b |g''(x)|^2 dx = ||\mathbf{D}_2\theta\Delta||^2$, where $\theta = (\theta_1, \ldots, \theta_J)$.

Optimization with the $\mathbf{D}_2$ penalty is equivalent to assuming a second order random-walk prior, that is, the improper prior $\mathbf{D}_2\theta \sim N_{J-2}(\mathbf{0}, \lambda^{-1}\mathbf{I}_{J-2})$. As λ becomes large, $g''(x)$ is forced toward zero and $g(x)$ becomes linear. Alternatively, a first order random walk prior is given by $\mathbf{D}_1\theta \sim N_{J-1}(\mathbf{0}, \lambda^{-1}\mathbf{I}_{J-1})$. When λ is large, adjacent basis functions are forced closer and $g'(x)$ is forced toward zero, yielding a constant $g(x)$.

The Bernstein polynomial is a special case of the B-spline with support $[0,1]$ (Petrone, 1999a,b). A Bernstein polynomial prior for a function g on $[0,1]$ is a discrete mixture of beta distributions with equispaced means and integer parameters; i.e. the functions $B_j(x)$ in (1) are

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

The resulting g is then transformed to $[0,b)$ ($b = \infty$ for some transformations) for use in baseline survival modeling (Gelfand and Mallick, 1995; Carlin and Hodges, 1999; Banerjee and Dey, 2005; Chang et al., 2005; Chen et al., 2014).

B-splines are now a standard tool for modeling hazard functions. Like the GP, the piecewise constant hazard is a special case, i.e. a first order B-spline with $d = 0$; piecewise exponental models have been used extensively in Bayesian survival analysis, e.g. Ibrahim et al. (2001). Existing approaches to modeling hazard functions using B-splines (Gray, 1992; Hennerfeind et al., 2006; Sharef et al., 2010) choose either equispaced knots over the spread of the observed data or knots at the empirical quantiles of the observed event times. Chen et al. (2014) and Li et al. (2015b) instead choose knot locations based on an approximation of underlying parametric family, e.g. S_θ indexed by θ.

2.3 Dirichlet process mixture model

A random probability measure G follows a DP (Ferguson, 1973) with parameters (α, G_0), where $\alpha > 0$ and G_0 is an appropriate probability measure defined on $\mathbb{R}^d$,

written as

$$G|\alpha, G_0 \sim DP(\alpha G_0), \tag{2}$$

if for any measurable nontrivial partition $\{B_l : 1 \leq l \leq k\}$ of $\mathbb{R}^d$, then the vector $(G(B_1), \ldots, G(B_k))'$ has a Dirichlet distribution with parameters $(\alpha G_0(B_1), \ldots, \alpha G_0(B_k))$. It follows that

$$G(B_l)|\alpha, G_0 \sim \text{Beta}(\alpha G_0(B_l), \alpha G_0(B_l^c)),$$

and therefore $E\{G(B_l)|\alpha, G_0\} = G_0(B_l)$ and $Var\{G(B_l)|\alpha, G_0\} = G_0(B_l)G_0(B_l^c)/(\alpha + 1)$. Thus G is centered at G_0 with precision α. The DP was used by Susarla and Van Ryzin (1976) to model and estimate the survival function for right-censored data; Müller et al. (2015) provide R code to implement this approach.

If $G|\alpha, G_0 \sim DP(\alpha G_0)$, then the process can be represented by the stick–breaking representation (Sethuraman, 1994),

$$G(\cdot) = \sum_{i=1}^{\infty} w_i \delta_\theta(\cdot), \tag{3}$$

where $\delta_\theta(\cdot)$ is Dirac measure at θ, $w_i = V_i \prod_{j<i}(1 - V_j)$, with $V_i|\alpha \overset{iid}{\sim} \text{Beta}(1, \alpha)$, and $\theta_i|G_0 \overset{iid}{\sim} G_0$. Note that $E(w_j) > E(w_{j+1})$ for all j, so the weights are stochastically ordered.

Convolving a DP with a parametric kernel, such as the normal, gives a DP mixture (DPM) model (Lo, 1984; Escobar and West, 1995). A simple DPM of Gaussian densities for continuous data $\varepsilon_1, \ldots, \varepsilon_n$ is given by

$$\varepsilon_i|G \overset{iid}{\sim} \int N(\mu, \sigma^2) dG(\mu, \sigma^2), \tag{4}$$

where $N(\mu, \sigma^2)$ denotes the normal density with mean μ and σ^2, and the mixing distribution, G, is a random probability measure defined on $\mathbb{R} \times \mathbb{R}^+$, following a DP. The stick–breaking representation recasts (4) as a countably infinite mixture of normals given by

$$\varepsilon_i|G \overset{iid}{\sim} \sum_{j=1}^{\infty} \left[V_j \prod_{k=1}^{j-1} (1 - V_k) \right] N(\mu_j, \sigma_j^2). \tag{5}$$

The prior distribution on ε_i is centered at the normal distribution; Griffin (2010) discusses prior specifications that control the "non–normalness" of this distribution.

2.4 Polya tree

A Polya tree (PT) successively partitions the reals $\mathbb{R}$ (or any other domain) into finer and finer partitions; each refinement of a partition doubles the number of partition sets by cutting the previous level's sets into two pieces; there are two sets at level 1, four sets at level 2, eight sets at 3, and so on. We focus on a PT centered at

the standard normal density, that is, $N(0,1)$ is the *centering distribution* for the Polya tree. At level j, the Polya tree partitions the real line into 2^j intervals $B_{j,k} = (\Phi^{-1}((k-1)2^{-j}), \Phi^{-1}(k2^{-j}))$ of probability 2^{-j} under Φ, $k = 1,\ldots,2^j$, where $\Phi(\cdot)$ is the cumulative distribution function of $N(0,1)$. Note that $B_{j,k} = B_{j+1,2k-1} \cap B_{j+1,2k}$. Given an observation ε is in set k at level j, i.e. $\varepsilon \in B_{j,k}$, it could then be in either of the two offspring sets $B_{j+1,2k-1}$ or $B_{j+1,2k}$ at level $j+1$. The conditional probabilities associated with these sets will be denoted by $Y_{j+1,2k-1}$ and $Y_{j+1,2k}$. Clearly they must sum to one, and so a common prior for either of these probabilities is a beta distribution (Ferguson, 1974; Lavine, 1992, 1994; Walker and Mallick, 1997, 1999; Hanson and Johnson, 2002; Hanson, 2006a; Zhao et al., 2009), given by

$$Y_{j,2k-1}|c \overset{ind.}{\sim} \text{Beta}(cj^2, cj^2), \quad j = 1,\ldots,J;\ k = 1,\ldots,2^{j-1},$$

where $c > 0$, which ensures that every realization of the process has a density, allowing the modeling of continuous data without the need of convolutions with continuous kernels.

The user–specified weight $c > 0$ controls how closely the posterior follows $N(0,1)$ in terms of L_1 distance (Hanson et al., 2008), with larger values forcing the PT process G closer to $N(0,1)$; often a prior is placed on c, e.g. $c \sim \Gamma(a,b)$. The PT is stopped at level J (typically $J = 5,6,7$); within the sets $\{B_{J,k} : k = 1,\ldots,2^J\}$ at the level J, G follows $N(0,1)$ (Hanson, 2006a). The resulting model for data $\varepsilon_1,\ldots,\varepsilon_n$ is given by

$$\varepsilon_i|G \overset{iid}{\sim} G, \tag{6}$$

where

$$G \sim PT_J(c, N(0,1)). \tag{7}$$

The corresponding density is given by

$$p(\varepsilon|\{Y_{j,k}\}) = 2^J \phi(\varepsilon) \prod_{j=1}^{J} Y_{j,\lceil 2^j \phi(\varepsilon)\rceil}, \tag{8}$$

where $\lceil\cdot\rceil$ is the ceiling function, and so a likelihood can be formed. For the simple model, the PT is conjugate. Let $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_n)$. Then

$$Y_{j,2k-1}|\varepsilon \overset{ind.}{\sim} \text{Beta}\left(cj^2 + \sum_{i=1}^{n} I\{\lceil 2^j\phi(\varepsilon_i)\rceil = 2k-1\}, cj^2 + \sum_{i=1}^{n} I\{\lceil 2^j\phi(\varepsilon_i)\rceil = 2k\}\right),$$

and $Y_{j,2k} = 1 - Y_{j,2k-1}$.

Location μ and spread σ parameters are melded with expression (6) and the PT prior (7) to make a median-μ location-scale family for data $y_1,\ldots,y_n$, given by

$$y_i = \mu + \sigma\varepsilon_i,$$

where the $\varepsilon_i \mid G \overset{iid}{\sim} G$ and G follows a PT prior as in expression (7), with the restriction $Y_{1,1} = Y_{1,2} = 0.5$. Allowing μ and σ to be random induces a mixture of Polya

trees (MPT) model for $y_1, \ldots, y_n$, smoothing out predictive inference (Lavine, 1992; Hanson and Johnson, 2002). Note that Jeffreys' prior under the normal model is a reasonable choice here (Berger and Guglielmi, 2001), and leads to a proper posterior (Hanson, 2006a).

3 Survival models

3.1 Proportional hazards

A proportional hazards (PH) model (Cox, 1972), for continuous data, is obtained by expressing the covariate-dependent survival function $S_{\mathbf{x}}(t)$ as

$$S_{\mathbf{x}}(t) = S_0(t)^{\exp(\mathbf{x}'\beta)}. \tag{9}$$

In terms of hazards, this model is

$$h_{\mathbf{x}}(t) = \exp(\mathbf{x}'\beta)h_0(t).$$

Note then that for two individuals with covariates $\mathbf{x}_1$ and $\mathbf{x}_2$, the ratio of hazard curves is constant and proportional to $\frac{h_{\mathbf{x}_1}(t)}{h_{\mathbf{x}_2}(t)} = \exp\{(\mathbf{x}_1 - \mathbf{x}_2)'\beta\}$, hence the name "proportional hazards." Cox (1972) is the second most cited statistical paper of all time (Ryan and Woodall, 2005), and the PH model is easily the most popular semiparametric survival model in statistics, to the point where medical researchers tend to compare different populations' survival in terms of instantaneous risk (hazard) rather than mean or median survival as in common regression models. Part of the popularity of the model has to do with the incredible momentum the model has gained from how easy it is to fit the model through partial likelihood (Cox, 1975) and its implementation in SAS in the procedure `PHREG`. The use of partial likelihood and subsequent counting process formulation (Andersen and Gill, 1982) of the model has allowed ready extension to stratified analyses, proportional intensity models, frailty models, and so on (Therneau and Grambsch, 2000).

The first Bayesian semiparametric approach to PH models posits a gamma process as a prior on the baseline cumulative hazard $H_0(t) = \int_0^t h_0(s)ds$ (Kalbfleisch, 1978); partial likelihood emerges as a limiting case (of the marginal likelihood as the precision parameter approaches zero). The use of the gamma process prior in PH models, as well as the beta process prior (Hjort, 1990), piecewise exponential priors, and correlated increments priors are covered in Ibrahim et al. (2001) (pp. 47–94) and Sinha and Dey (1997). Other approaches include what are essentially Bernstein polynomials (Gelfand and Mallick, 1995; Carlin and Hodges, 1999) and penalized B–splines (Hennerfeind et al., 2006; Kneib and Fahrmeir, 2007). The last two models are available in the free software BayesX (Belitz et al., 2015) which can be called from R via the packages `R2BayesX` and `BayesX` (Umlauf et al., 2015). The BayesX functions allow for a general additive (including partially linear) PH

model to be easily fit, including time-dependent covariates; BayesX also accommodates spatial frailties, discussed in Section 4.1. PH models with Polya tree baselines were considered by Hanson (2006a), Hanson and Yang (2007), Zhao et al. (2009), and Hanson et al. (2009) and can be fit in the `SpBayesSurv` package for R.

Stratified PH model posits a separate hazard function across levels of strata $s = 1,\ldots,S$,

$$h_{\mathbf{x},s}(t) = \exp(\mathbf{x}'\beta)h_{0s}(t).$$

A version of this model based on Bernstein polynomials is given by Carlin and Hodges (1999); B-splines were considered by Cai and Meyer (2011). The stratified PH model can also be fit using SAS `PHREG` assuming piecewise exponential priors, i.e. piecewise constant baseline hazard functions. A version of the stratified model that SAS fits, but with a "Polya tree" type prior on the hazard was considered by Dukić and Dignam (2007). Note that BayesX can also fit stratified models based on B-splines by including a time-varying regression effect for the categorical strata variable.

3.2 Accelerated failure time

An accelerated failure time (AFT) model is obtained by expressing the covariate–dependent survival function $S_{\mathbf{x}}(t)$ as

$$S_{\mathbf{x}}(t) = S_0\{\exp(-\mathbf{x}'\beta)t\}. \tag{10}$$

This is equivalent to the linear model for the log transformation of the corresponding time-to-event response variable, T,

$$\log T = \mathbf{x}'\beta + \varepsilon, \tag{11}$$

where $\exp(\varepsilon) \sim S_0$. The mean, median, and any quantile of survival for an individual with covariates $\mathbf{x}_1$ is changed by a factor of $\exp\{(\mathbf{x}_1 - \mathbf{x}_2)'\beta\}$ relative to those with covariates $\mathbf{x}_2$.

An early frequentist least-squares teatment of the AFT model with right-censored data is due to Buckley and James (1979); the Buckley-James estimator is implemented in Frank Harrell's `Design` library for R (Alzola and Harrell, 2006). The R packages `emplik` and `bujar` have various extensions. More refined estimators followed in the 1990's (Ying et al., 1995; Yang, 1999) focusing on median-regression.

From a Bayesian nonparametric perspective, the first approach, based on a Dirichlet process prior, obtained approximate marginal inferences to the AFT model (Christensen and Johnson, 1988); a full Bayesian treatment using the Dirichlet process is not practically possible (Johnson and Christensen, 1989). Approaches based on Dirichlet process mixture models have been considered by Kuo and Mallick (1997), Kottas and Gelfand (2001) and Hanson (2006b). Dirichlet process mixtures

"fix" the discrete nature of the Dirichlet process, as do other discrete mixtures of continuous kernels. We refer the reader to Komárek and Lesaffre (2007) for an alternative approach based on finite mixtures of normal distributions, and Komárek and Lesaffre (2008) based on an approximating B-spline, both available in the R package `bayesSurv`. Polya tree priors that have continuous densities can directly model the distribution of ε in expression (11) (Walker and Mallick, 1999; Hanson and Johnson, 2002; Hanson, 2006a; Hanson and Yang, 2007; Zhao et al., 2009). AFT models with Polya tree baseline densities can be fit in the `spBayesSurv` package for R.

Although PH is by far the most commonly-used semiparametric survival model, several studies have shown vastly superior fit and interpretation from AFT models (Hanson and Yang, 2007; Hanson, 2006a; Kay and Kinnersley, 2002; Orbe et al., 2002; Hutton and Monaghan, 2002). Cox pointed out himself (Reid, 1994) *"... the physical or substantive basis for ... proportional hazards models ... is one of its weaknesses ... accelerated failure time models are in many ways more appealing because of their quite direct physical interpretation ..."*. However, similar to the PH model, standard AFT models also impose constraints so that survival curves from different covariate levels are not allowed to cross, which is unrealistic in many practical applications (e.g., De Iorio et al., 2009). For these data that do not follow AFT assumptions, we next discuss two generalizations of the AFT model that allow for crossing survival and hazards curves. The two approaches are the *linear dependent Dirichlet process mixture*, which can be interpreted as a mixture of parametric AFT models, and the *linear dependent tailfree process*, which is an AFT model with very general baseline functions that are covariate-dependent. Both augmentations are examples of "density regressions," allowing the entire survival density $f_{\mathbf{x}}(t)$ to change smoothly with covariates $\mathbf{x}$.

3.2.1 Linear dependent Dirichlet process

By considering a Dirichlet process mixture of normal distributions for the errors in (11) (Kuo and Mallick, 1997), the distribution for the log survival time is the distribution of ε_i, given by (5), shifted by the linear predictor $\eta_i = \mathbf{x}_i'\beta$. Specifically,

$$y_i|\beta, G \overset{ind.}{\sim} \sum_{j=1}^{\infty} w_j N(\mu_j + \mathbf{x}_i'\beta, \sigma_j^2),$$

where $G(\cdot) = \sum_{j=1}^{\infty} w_j \delta_{(\mu_j,\sigma_j^2)}(\cdot)$ is a Dirichlet process. The interpretation of the components of β are as usual and the model can be fit using standard algorithms for Dirichlet process mixture models (Neal, 2000).

The linear dependent Dirichlet process mixture (LDDPM) (De Iorio et al., 2009; Jara et al., 2010, 2011; Zhou et al., 2015b) can be interpreted as a generalization of the previous model, which arises by additionally mixing over the regression coefficients, yielding a mixture of log-normal AFT models. Specifically, the LDDPM model is given by

$$y_i|G \overset{ind.}{\sim} \sum_{j=1}^{\infty} w_j N(\mathbf{x}_i'\beta_j, \sigma_j^2), \tag{12}$$

where $\mathbf{x}_i$ now includes a '1' for the intercept, $w_i = V_i \prod_{j<i}(1-V_j)$, with $V_i|\alpha \overset{iid}{\sim}$ Beta$(1,\alpha)$, and $\beta_j \overset{iid}{\sim} N(\mathbf{m}_0, \mathbf{V}_0)$ and $\sigma_j^{-2} \overset{iid}{\sim} \Gamma(a_0, b_0)$.

The model trades easy interpretability offered by a single β for greatly increased flexibility. In particular, the LDDPM model does not stochastically order survival curves from different predictors $\mathbf{x}_{i_1}$ and $\mathbf{x}_{i_2}$, and both the survival and hazard curves can cross.

3.2.2 Linear dependent tailfree process

A Polya trees defines the conditional probabilities $Y_{j+1,2k-1}$ and $Y_{j+1,2k}$ as beta distributions. However, one can instead define a logistic regression for each of these probabilities, allowing the *entire* shape of the density to change with covariates; this is the approach considered by Jara and Hanson (2011). Given covariates $\mathbf{x}$, the linear dependent tailfree process (LDTFP) models $(Y_{j+1,2k-1}, Y_{j+1,2k})$ through logistic regressions

$$\log\{Y_{j+1,2k-1}(\mathbf{x})/Y_{j+1,2k}(\mathbf{x})\} = \mathbf{x}'\tau_{j,k},$$

where $\mathbf{x}$ includes an intercept. There are $2^J - 1$ regression coefficient vectors $\tau = \{\tau_{j,k}\}$; e.g. for $J = 3$, $\{\tau_{0,1}, \tau_{1,1}, \tau_{1,2}, \tau_{2,1}, \tau_{2,2}, \tau_{2,3}, \tau_{2,4}\}$. Let $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_n]'$ be the $n \times p$ design matrix. Following Jara and Hanson (2011), each is assigned an independent normal prior, $\tau_{j,k} \sim N_p\left(\mathbf{0}, \frac{2}{c(j+1)^2}\Psi\right)$. Jara and Hanson (2011) discussed the case $\Psi = n(\mathbf{X}'\mathbf{X})^{-1}$, generating a g-prior Zellner (1983) for the tailfree regression coefficients. By setting $\tau_{0,1} \equiv \mathbf{0}$, the resulting LDTFP is almost surely a median-zero probability measure for every $\mathbf{x} \in \mathscr{X}$, important to avoid identifiability issues.

Augmenting (8), the random density is given by

$$g_{\mathbf{x}}(\varepsilon) = \phi(\varepsilon) 2^J \prod_{j=1}^{J} Y_{j,\lceil 2^j \Phi(\varepsilon)\rceil}(\mathbf{x}).$$

Since the $\{Y_{j,k}\}$ are modeled with logistic-normal distribution instead of beta, the resulting random density is a tailfree process. The final AFT model with LDTFP baseline is given by

$$y_i = \mathbf{x}_i'\beta + \sigma\varepsilon_i, \;\; \varepsilon_i|\tau \overset{ind.}{\sim} g_{\mathbf{x}_i}. \tag{13}$$

Unlike the LDDPM, the LDTFP separates survival into one distinct trend $\mathbf{x}'\beta$ and an evolving log-baseline survival density $g_{\mathbf{x}}$. By forcing $g_{\mathbf{x}}$ to be median-zero, e^{β_j} gives a factor by how median survival changes when x_j is increased *just as in standard AFT models*. This heightened interpretability in terms of median-regression in the presence of heteroscedastic error allows a fit of the LDTFP model to easily relate covariates $\mathbf{x}$ to median survival.

The LDTFP models the probability of falling above or below quantiles of the $N(\mathbf{x}'\beta, \sigma^2)$ distribution, but in terms of conditional probabilities. This model can be viewed as a particular kind of quantile regression model. Koenker and Hallock (2001) suggest that "*...instead of estimating linear conditional quantile models, we could instead estimate a family of binary response models for the probability that the response variable exceeded some prespecified cutoff values.*" However, Koenker and Hallock (2001) prefer the linear (in covariates) quantile specification because "*...it nests within it the iid error location shift model of classical linear regression.*" By augmenting a median-zero tailfree process with a general trend $\mathbf{x}'\beta$ we accomplish the same objective, nesting the ubiquitous normal-errors linear model within a highly flexible median regression model, but with heteroscedastic error that changes shape with covariate levels $\mathbf{x} \in \mathscr{X}$.

Both the LDDPM and the LDTFP model the entire density at every covariate level $\mathbf{x} \in \mathscr{X}$, so full density and hazard estimates are available, accompanied by reliable interval estimates, unlike many median (and other quantile) regression models. Both models are implemented as user-friendly functions calling compiled FORTRAN in `DPpackage` or calling compiled C++ in `spBayesSurv` for `R`. These functions accommodate general interval-censored data (including current status data); the latter package also allows for spatial correlation. If only a trend function is desired one could instead use quantile regression models, such as the ones implemented in the excellent `quantreg` package in `R` (Koenker, 2008).

3.3 Proportional odds

The proportional odds (PO) model has recently gained attention as an alternative to the PH and AFT models. PO defines the survival function $S_{\mathbf{x}}(t)$ for an individual with covariate vector $\mathbf{x}$ through the relation

$$\frac{S_{\mathbf{x}}(t)}{1-S_{\mathbf{x}}(t)} = \exp\{-\mathbf{x}'\beta\}\left(\frac{S_0(t)}{1-S_0(t)}\right). \tag{14}$$

The odds of dying before any time t are $\exp\{(\mathbf{x}_1-\mathbf{x}_2)'\beta\}$ times greater for those with covariates $\mathbf{x}_1$ versus $\mathbf{x}_2$.

The first semiparametric approaches to PO models involving covariates are due to Cheng et al. (1995), Murphy et al. (1997), and Yang and Prentice (1999). A semiparametric frequentist implementation of the PO model is available in the package `timereg` (Martinussen and Scheike, 2006) for `R`. Bayesian nonparametric approaches for the PO model have been based on Bernstein polynomials (Banerjee and Dey, 2005), B-splines (Wang and Dunson, 2011; Lin and Wang, 2011), and Polya trees (Hanson, 2006a; Hanson and Yang, 2007; Zhao et al., 2009; Hanson et al., 2011).

The PH, AFT, and PO models all make overarching assumptions about the data generating mechanism for the sake of obtaining succinct data summaries. An impor-

tant aspect associated with the Bayesian nonparametric formulation of these models is that, by assuming the *same, flexible model* for the baseline survival function, they are placed on a common ground (Hanson, 2006a; Hanson and Yang, 2007; Zhang and Davidian, 2008; Zhao et al., 2009; Hanson et al., 2011). Furthermore, parametric models are special cases of the nonparametric models. Differences in fit and/or predictive performance can therefore be attributed to the *survival* models only, rather than to additional possible differences in quite different nonparametric models or estimation methods.

Of the Bayesian approaches based on Polya trees considered by Hanson (2006a), Hanson and Yang (2007), Zhao et al. (2009) and Hanson et al. (2011), the PO model was chosen over PH and AFT according to the log-pseudo marginal likelihood (LPML) criterion (Geisser and Eddy, 1979). In three of these works, the parametric log-logistic model, a special case of PO that also has the AFT property, was chosen. This may be due to the fact that the PO assumption implies that hazard ratios $\lim_{t\to\infty}\frac{h_{\mathbf{x}_1}(t)}{h_{\mathbf{x}_2}(t)}=1$, that is, eventually everyone has the same risk of dying tomorrow. These authors also found that, everything else being equal, the actual semiparametric model chosen (PO, PH or AFT) affects prediction far more than whether the baseline is modeled nonparametrically. It is worth noting that none of these papers favored the semiparametric PH model in actual applications.

3.4 Other semiparametric models

PH, AFT, and PO are three of many semiparametric survival models used in practice. There are a few more hazard-based models including the additive hazards (AH) model (Aalen, 1980, 1989), given by

$$h_{\mathbf{x}}(t)=h_0(t)+\mathbf{x}'\boldsymbol{\beta},$$

which is implemented in the `timereg` package for `R`. An empirical Bayes approach to this model based on the gamma process was implemented by Sinha et al. (2009). Fully Bayesian approaches require an elaborate model specification to incorporate the rather awkward constraint $h_0(t)+\mathbf{x}'\boldsymbol{\beta}\geq 0$ for $t>0$ (Yin and Ibrahim, 2005; Dunson and Herring, 2005). Recently, there has been some interest in the accelerated hazards model (Chen and Wang, 2000; Zhang et al., 2011; Chen et al., 2014), given by

$$h_{\mathbf{x}}(t)=h_0\{\exp(-\mathbf{x}'\boldsymbol{\beta})t\}.$$

This model allows hazard and survival curves to cross.

Finally, several interesting "super models" have been proposed in the literature, including non-proportional hazard regression models that include PH as a special case (Devarajan and Ebrahimi, 2011), generalized odds-rate hazards models that include PH and PO as special cases (Dabrowska and Doksum, 1988; Scharfstein et al., 1998), Box-Cox transformation regression models that include PH and AH as spe-

cial cases (Yin and Ibrahim, 2005; Martinussen and Scheike, 2006), and extended hazard regression models that include both PH and AFT as special cases (Chen and Jewell, 2001; Li et al., 2015b).

4 Spatial dependence

When survival data are spatially correlated, it is often of scientific interest to investigate possible spatial dependence in survival outcomes after adjusting for known subject-specific covariate effects. Such spatial dependence is often due to region-specific similarities in ecological and/or social environments that are typically not measurable. We next discuss two general approaches, *frailty* and *copula*, for incorporating spatial dependence into the semiparametric models presented in Section 3, followed by some other possibilities.

4.1 Spatial frailty modeling

Frailties have been frequently used to induce correlation among related survival times in models which have a linear predictor. The linear predictor is augmented $\eta_i = \mathbf{x}_i'\beta + v_i$, where v_i is a random effect, termed "frailty", accounting for heterogeneity after adjusting for covariates. So-called shared frailty models have one common random effect within each group, e.g. $v_i = z_{g_i}$ where $g_i \in \{1,\ldots,G\}$ is the group – e.g. county, hospital, family – to which observation i belongs. Early literature considered exchangeable frailties with $z_1,\ldots,z_G \overset{iid}{\sim} H$, where H was constrained to be mean or median zero to avoid confounding with the baseline function.

In the case of spatial survival data, one can extend the frailty model by including a spatial effect, e.g.,

$$\eta_i = \mathbf{x}_i'\beta + \gamma_i,\ \gamma_i = v_i + w_i,$$

where the frailty term γ_i incorporates the effects of both heterogeneity (via the non-spatial frailty v_i) and spatial dependence (through the spatial frailty w_i); note that, however, in applications often only spatial dependence is modeled ($\gamma_i = w_i$) or exchangeable dependence ($\gamma_i = v_i$). Spatial frailty models have been widely discussed in the literature and correspond to particular cases of hierarchical models. Such models are usually grouped into two general settings according to their underlying data structure: *point-referenced* (geostatistical) data, where the location $\mathbf{s}_i$ varies continuously throughout a fixed study region $\mathscr{D}$, and *areal* (lattice) data, where the study region is partitioned into a finite number of areal units with well-defined boundaries (Banerjee et al., 2014).

4.1.1 Point-referenced data modeling

In modeling point-referenced data, the non-spatial frailty term v_i is often specified $v_i \overset{iid}{\sim} N(0,\sigma^2)$, and the spatially correlated frailties $\mathbf{w} = (w_1,\ldots,w_n)$ can be specified to have a multivariate Gaussian distribution:

$$\mathbf{w} \sim N_n(\mathbf{0}, \theta^2\mathbf{R}), \tag{15}$$

where N_n denotes the n-dimensional Gaussian distribution, θ^2 measures the amount of spatial variation across locations, and the (i,j)th element of $\mathbf{R}$, denoted by $\mathbf{R}_{ij}$, is the correlation between w_i and w_j. An isotropic correlation function is commonly used to construct $\mathbf{R}$, where the correlation of any two subjects is a function solely of the distance d_{ij} between their locations $\mathbf{s}_i$ and $\mathbf{s}_j$, i.e., $\mathbf{R}_{ij} = \rho(d_{ij})$. A flexible, frequently-used correlation function is the Matérn

$$\rho(d_{ij}) = \frac{(\phi d_{ij})^\nu K_\nu(\phi d_{ij})}{2^{\nu-1}\Gamma(\nu)}, \tag{16}$$

where K_ν is a modified Bessel function of the third kind, $\phi > 0$ measures the spatial decay over distance, and $\nu > 0$ is a parameter controlling the smoothness of the realized random field. Interested readers are referred to Banerjee et al. (2014) for further discussion of correlation functions. Note that the Matérn reduces to the exponential $\rho(d_{ij}) = \exp(-\phi d_{ij})$ for $\nu = 0.5$ and the Gaussian $\rho(d_{ij}) = \exp(-\phi^2 d_{ij}^2)$ when $\nu \to \infty$. Under the above prior specifications of exchangeable normal v_i and spatially correlated w_i, the resulting multivariate Gaussian distribution on frailties $\gamma = (\gamma_1,\ldots,\gamma_n)$ is

$$\gamma \sim N_n\left\{\mathbf{0}, \theta^2\mathbf{R} + \sigma^2\mathbf{I}\right\}. \tag{17}$$

With this representation, the non-spatial effect variance σ^2 is often called the *nugget*, the spatial effect variance θ^2 is called the *partial sill*, and the total effect variance $\theta^2+\sigma^2$ is called the *sill*. The rationale of including the nugget effect is that we don't expect all remaining individual heterogeneity to be accounted for by the spatial story, as other factors (e.g., measurement error, replication error, micro-scale error) may also potentially explain the heterogeneity. In Henderson et al. (2002), the term $\tau = \theta^2/(\theta^2+\sigma^2)$ is called the *nugget effect* and interpreted as the proportion of the heterogeneity variance that is explained by spatial effects.

For posterior inference, MCMC requires computing the inverse and determinant of n-dimensional correlation matrix $\mathbf{R}$ in each iteration. With an increasing sample size n, such computation becomes very expensive and even unstable due to a large amount of numerical operations. This situation is often referred to as "the big n problem." Various approaches have been developed to approximate the correlation function such as predictive process models (Banerjee et al., 2008; Finley et al., 2009), sparse approximations (Furrer et al., 2006; Kaufman et al., 2008), and the full scale approximation (FSA) method (Sang and Huang, 2012). The last approximation is the summation of the former two approximations, which can capture both large- and small-scale spatial dependence. The FSA has been successfully applied

to model point-referenced survival data in Zhou et al. (2015b) and implemented in the R package `spBayesSurv`.

4.1.2 Areal data modeling

In the case of areal data, the whole study region $\mathscr{D}$ is often partitioned into a finite number of areas, say $B_1,\ldots,B_G$, and a common frailty is assumed for the subjects within each area, i.e.

$$\eta_i = \mathbf{x}_i'\beta + \gamma_{g_i},\ \gamma_j = v_j + w_j, j = 1,\ldots,G.$$

Here the non-spatial frailty v_j for each area is typically assigned a mean-zero normal distribution with variance σ^2. For the spatial frailty term w_j, there has been two general approaches. First, one can assume a fully-specified mean-zero multivariate Gaussian distribution on $\mathbf{w} = (w_1,\ldots,w_G)$ with covariance matrix $\theta^2\mathbf{R}$, where $\mathbf{R}_{ij}$ is modeled using a traditional correlation function like the Matérn in (16) but with d_{ij} representing the distance between two areal centroids. Another way is to consider an intrinsic conditionally autoregressive (ICAR) model. Let $a_{ij} = 1$ if areas B_i and B_j share a nontrivial border (i.e. a connected curve in $\mathbb{R}^2$ that is more than one point) and $a_{ij} = 0$ otherwise; set $a_{ii} = 0$. Then the $G \times G$ matrix $\mathbf{A} = [a_{ij}]$ is called the adjacency matrix for the region $\mathscr{D}$. The ICAR prior is defined through the set of all conditional distributions

$$w_j|\{w_i : i \neq j\} \sim N\left(\bar{w}_j,\ \theta^2/a_{j+}\right),\ j = 1,\ldots,G, \tag{18}$$

denoted $\mathbf{w} \sim \text{ICAR}(1/\theta^2)$, where a_{j+} is the number of neighbors of area B_j, $\bar{w}_j = \frac{1}{a_{j+}}\sum_{i:a_{ij}=1} w_i$ is the sample mean of the a_{j+} values of the neighboring areal unit frailties, and θ^2/a_{j+} is the conditional variance. Note that the ICAR model induces an improper joint density, and the constraint $\sum_{j=1}^G w_j = 0$ is commonly used to avoid identifiability issues. Another common fix is to assume a proper CAR model by multiplying the conditional mean $\bar{w}_i$ in (18) by a shrinkage scale parameter ρ, where $0 \leq \rho < 1$; it is generally difficult to estimate ρ and θ^2 simultaneously.

4.1.3 Related literature

Henderson et al. (2002) modeled the spatial structure of leukemia survival data using both district-level and point-referenced frailty effects in the context of the PH model. In their point-referenced analysis, a multivariate gamma distribution for $(e^{\gamma_1},\ldots,e^{\gamma_n})'$ was constructed so that each marginal has a gamma distribution with mean 1 and variance $\sigma^2 + \theta^2$, and the correlation between e^{γ_i} and e^{γ_j} takes the form defined in (17). In their district-level analysis, they considered a linear predictor with individual frailties as $\eta_i = \mathbf{x}_i'\beta + \gamma_i$, where $e^{\gamma_i}|\mu_{g_i} \sim \Gamma\left(1/\xi, 1/(\xi\mu_{g_i})\right)$. They then assumed a multivariate Gaussian distribution on the latent effects $\mu = (\mu_1,\ldots,\mu_G)$

with the correlation function between the ith and jth district modeled via the powered exponential and Matérn. They also considered the ICAR specification on μ and found that the multivariate Gaussian via a Matérn correlation with $\nu = 2$ had the best fit based on the DIC goodness-of-fit criterion.

Pan et al. (2014) fitted the semiparametric PH model with ICAR frailties to interval censored data with the baseline hazard function modeled via B-splines. Lin et al. (2015) duplicated this model without the ICAR frailties. Using the same methodology, a special case of interval-censored data, current-status data, was presented in Cai et al. (2011). The aforementioned models can be fit in the `ICBayes` R package. Li and Ryan (2002) modeled the district-level frailty effect using a fully-specified multivariate normal prior within the framework of PH, and applied the model to detect prognostic factors leading to childhood asthma. All of these approaches are essentially a special case of the general models previously presented in Kneib (2006) and Hennerfeind et al. (2006), which can be efficiently fit in the freely-available program BayesX or the R package `R2BayesX`; the latter package uses compiled code and places the B-spline prior on the log-hazard instead of the hazard. An advantage of the models fitted in BayesX is that both areal and point-referenced data are accommodated as well as nonparametric additive effects. In addition, the R package `spatsurv` can also fit the PH model with multivariate Gaussian frailties, where the baseline hazard is modeled either parametrically or nonparametrically via B-splines.

Banerjee and Carlin (2003) developed a semiparametric PH frailty model for capturing spatio-temporal heterogeneity in survival of women diagnosed with breast cancer in Iowa, using a mixture of beta densities baseline. Banerjee et al. (2003) applied the Weibull parametric PH frailty model to infant mortality data in Minnesota, where the county-level frailties were assumed to have either an uncorrelated zero-mean Gaussian prior, an ICAR prior, or a fully-specified multivariate Gaussian prior as in (15). They showed that the fully specified prior provides the best model fitting in terms of DIC in the analysis of the infant data. Banerjee and Dey (2005) utilized the same frailty modeling technique for capturing spatial heterogeneity within the framework of semiparametric PO, found that the proper CAR prior yielded the best fit in the application to a subset of SEER breast cancer data. Zhao et al. (2009) considered either an AFT, PH or PO model with ICAR frailties, where the baseline function was assumed to have a mixture of Polya trees prior. Zhang and Lawson (2011) and Wang et al. (2012) developed parametric and semiparametric AFT models with ICAR frailties, respectively. Chernoukhov (2013) extended the additive hazards model for allowing various spatial dependence structures in his dissertation. Zhou et al. (2015c) extended the generalized model in (13) by allowing frailties accommodating spatial correlation via the ICAR prior distribution. The models proposed in Zhao et al. (2009) and Zhou et al. (2015c) can be fit in the R package `spBayesSurv`. Other references focusing on spatial frailty modeling and its application include McKinley (2007), Diva et al. (2008), Darmofal (2009), Liu (2012), Ojiambo and Kang (2013), Dasgupta et al. (2014), Li et al. (2015a), and among others.

4.2 Spatial copula modeling

Spatial copulas are just beginning to become popular in geostatistics. The use of copulas in the spatial context was first proposed by Bárdossy (2006), where the empirical variogram is replaced by empirical copulas to investigate the spatial dependence structure. The spatial copula approach offers an appealing way to separate modeling from the spatial dependence structure for multivariate distributions. Copulas completely describe association among random variables separately from their univariate distributions and thus capture joint dependence without the influence of the marginal distribution (Li, 2010). In the context of survival models, the idea of spatial copula approach is to first assume that the survival time T_i at location $\mathbf{s}_i$ marginally follows a model $S_{\mathbf{x}_i}(t)$ introduced in Section 3, then model the joint distribution of $(T_1,\ldots,T_n)'$ as

$$F(t_1,\ldots,t_n) = C(F_{\mathbf{x}_1}(t_1),\ldots,F_{\mathbf{x}_n}(t_n)), \tag{19}$$

where $F_{\mathbf{x}_i}(t) = 1 - S_{\mathbf{x}_i}(t)$ is the cumulative distribution function and the function C is an n-copula used to capture spatial dependence. If we let $U_i = F_{\mathbf{x}_i}(T_i)$, then the problem is reduced to constructing a copula for modeling the joint distribution of $\mathbf{U} = (U_1,\ldots,U_n)$. Hereafter we assume that U_i follows a uniform distribution on $[0,1]$ for all locations $\mathbf{s}_i$; i.e. the survival model $S_{\mathbf{x}_i}(t)$ is assumed to be correctly specified. In fact, copulas are all the joint cumulative distribution functions on the unit hypercube with uniform marginal distributions. We refer interested readers to Nelsen (2006) for general introduction to copulas and to Smith (2013) for Bayesian approaches to copula modeling.

In the geostatistical framework, the multivariate spatial copula of $\mathbf{U}$ is often constructed so that for any selected two locations $\mathbf{s}_i$ and $\mathbf{s}_j$, the bivariate copula (i.e., joint distribution) of (U_i,U_j) does not depend on the locations $\mathbf{s}_i$ and $\mathbf{s}_j$ but on their distance d_{ij} only. However, such construction is not a trivial task. Here we introduce a spatial version of the Gaussian copula and refer readers to Li (2010) for further discussion of other theoretical spatial copulas. Define $Z_i = \Phi^{-1}\{U_i\}$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function, then we have $Z_i \sim N(0,1)$ for all i. If we further assume that $\mathbf{Z} = (Z_1,\ldots,Z_n)'$ follows a multivariate normal distribution with mean zero and covariance $\mathbf{R}$, i.e., $\mathbf{Z} \sim N_n(\mathbf{0},\mathbf{R})$, then the induced joint distribution of $\mathbf{U}$ is called the Gaussian copula, which is given by

$$C(u_1,\ldots,u_n) = \Phi_n\left(\Phi^{-1}\{u_1\},\ldots,\Phi^{-1}\{u_n\};\mathbf{R}\right), \tag{20}$$

where $\Phi_n(\ldots;\mathbf{R})$ denotes the distribution function of $N_n(\mathbf{0},\mathbf{R})$. Note that all the diagonal elements of $\mathbf{R}$ are ones, so we refer to $\mathbf{R}$ as the correlation matrix thereafter. The Gaussian copula has a symmetrical density, which can be written as

$$c(u_1,\ldots,u_n) = |\mathbf{R}|^{-1/2}\exp\left\{\frac{1}{2}\mathbf{z}'(\mathbf{R}^{-1}-\mathbf{I})\mathbf{z}\right\}, \tag{21}$$

where $\mathbf{z} = (z_i, \ldots, z_n)'$ with $z_i = \Phi^{-1}\{u_i\}$ and $\mathbf{I}$ is the identity matrix. The spatial dependence structure of the Gaussian copula is induced by constructing the correlation matrix $\mathbf{R}$ using classical geostatistical models. For example, the (i, j)th element of $\mathbf{R}$ can be defined using the Matérn in (16) with a nugget effect τ, that is, $\mathbf{R}_{ij} = \tau\rho(d_{ij})$ for $i \neq j$, where $1 < \tau < 1$. Under the spatial Gaussian copula, the joint density of $(T_1, \ldots, T_n)$ takes the form

$$f(t_1, \ldots, t_n) = |\mathbf{R}|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{R}^{-1} - \mathbf{I}_n)\mathbf{z}\right\} \prod_{i=1}^{n} f_{\mathbf{x}_i}(t_i), \tag{22}$$

where $z_i = \Phi^{-1}\{F_{\mathbf{x}_i}(t_i)\}$ and $f_{\mathbf{x}_i}(t_i)$ is the density function of T_i. The use of spatial copulas has not been widely applied for modeling survival data that are subject to spatial correlation. Li and Lin (2006) successfully applied the spatial Gaussian copula approach to a semiparametric PH model and proposed spatial semiparametric estimating equations that yield consistent and asymptotically normal estimators. Zhou et al. (2015b) considered the LDDPM marginal model given in (12) using the same Gaussian copula for capturing spatial dependence structure, where MCMC algorithms were used to obtain posterior inferences. Zhou et al. (2015b) also provided a Bayesian version of the model considered in Li and Lin (2006) using piecewise exponential baseline specifications. The R package `spBayesSurv` can fit the aforementioned copula-based Bayesian survival models.

The spatial Gaussian copula approach can also be extended for fitting lattice data, for which constructing the correlation matrix $\mathbf{R}$ of $\mathbf{Z} = (Z_1, \ldots, Z_n)$ becomes a challenging task. One may consider a random effects model for $\mathbf{Z}$ based on the partition of the domain $\mathscr{D}$ into G districts, that is,

$$Z_i = \mu_{g_i} + \varepsilon_i,\ \mu \sim N_G(\mathbf{0}, \mathbf{B}\Omega\mathbf{B}),\ \varepsilon_i \overset{ind}{\sim} N\left(0, \frac{\sigma^2}{\omega_{g_i g_i} + \sigma^2}\right), g_i \in \{1, \ldots, G\}, \tag{23}$$

where $\mu = (\mu_1, \ldots, \mu_G)'$ are the random effects, $\Omega = [\omega_{ij}]$ is a $G \times G$ matrix introducing spatial dependence to μ, $\mathbf{B} = \text{diag}\left(1/\sqrt{\omega_{11} + \sigma^2}, \ldots, 1/\sqrt{\omega_{GG} + \sigma^2}\right)$, and ε_i is the error term independent of the spatial random effects. Note that $Var(Z_i) = 1$. Popular models for Ω include multivariate Gaussian coupled with a spatial covariance function, ICAR, proper CAR and many others. Li et al. (2015b) derived the implied correlation matrix $\mathbf{R} = cov(\mathbf{Z})$ under the ICAR model, which only involves one unknown quantity ψ^*. A smaller value of ψ^* corresponds to stronger spatial dependence. With the specification of $\mathbf{R}$, one can model joint cumulative distribution function of $(T_1, \ldots, T_n)$ by

$$F(t_1, \ldots, t_n) = \Phi_n\left(\Phi^{-1}\{F_{\mathbf{x}_i}(t_i)\}, \ldots, \Phi^{-1}\{F_{\mathbf{x}_n}(t_n)\}; \mathbf{R}\right). \tag{24}$$

4.3 Other spatial dependence modelings

Zhao and Hanson (2011) considered a stratified PH model:

$$S_{\mathbf{x}_i} = S_{0g_i}(t)^{\exp(\mathbf{x}_i'\beta)},\ g_i \in \{1,\ldots,G\},$$

where each region-specific baseline $S_{0j}(\cdot)$ approximately follows a mixture of Polya trees prior centered at a parametric log-logistic family. The spatial dependence among the $\{S_{01}(\cdot),\ldots,S_{0G}(\cdot)\}$ is induced through proper CAR priors on the logit transformed Polya tree conditional probabilities $\{Y_{l,k}\}$. Hanson et al. (2012) extended this idea to fit a Bayesian semiparametric temporally-stratified PH model with spatial frailties. Stratified AFT models with ICAR areal frailties are considered by Zhou et al. (2015c).

In modeling areal data, spatial dependence is often due to unadjusted district-level risk factors that may potentially relate to survival outcomes. Zhao and Hanson (2011) note that spatial frailties serve as proxies to unmeasured region-level covariates, but are less-precise adjustments since region-level covariates (such as shortest distance to a clinic) are unlikely to sharply change at areal boundaries. Therefore it is natural to introduce spatial dependence by allowing frailties to depend on region-level covariates, especially when information is available on each region that may affect the survival outcome beyond the recorded covariates. For this reason, Zhou et al. (2015a) proposed a region-level covariate adjusted frailty PH model. Specifically, with the linear predictor $\eta_i = \mathbf{x}_i'\beta + \gamma_{g_i}$, they assume an LDTFP prior on the frailties, i.e., $\gamma_j|\mathbf{z}_j \sim g_{\mathbf{z}_j}(\cdot)$, where $\mathbf{z}_j$ is a vector of region-level covariates. This model can be fit in the `DPpackage` for R.

5 Illustrations

Both of the frailty and copula modeling approaches are illustrated using real-life datasets. All the analyses are implemented using the R packages `spBayesSurv`. The fitted models are compared in terms of the log pseudo marginal likelihood (LPML) developed by Geisser and Eddy (1979). Note that the frailties used in frailty models are either exchangeable v_i or spatial w_i, but not both $v_i + w_i$.

5.1 SEER cancer data

The Surveillance Epidemiology and End Results (SEER) program of the National Cancer Institute (`seer.cancer.gov`) is an authoritative source of information on cancer incidence and survival in the US, providing county-level cancer data on an annual basis for particular states for public use. Areal-referenced SEER data have been analyzed by many authors in the context of spatial frailty models (e.g.,

Banerjee and Carlin, 2003; Banerjee and Dey, 2005; Zhao et al., 2009; Zhao and Hanson, 2011; Wang et al., 2012; Zhou et al., 2015a,c).

For illustration, we analyze a subset of the Iowa SEER breast cancer survival data, which consists of a cohort of 1073 Iowan women, who were diagnosed with malignant breast cancer starting in 1995, and enrollment and follow-up continued through the end of 1998. This data set has been analyzed in Zhao et al. (2009), and Zhou et al. (2015a). The observed survival time, from 1 to 48, is defined as the number of months from diagnosis to either death or the last follow-up. Here we assume that only deaths due to metastasis of cancerous nodes in the breast are events, while the deaths from other causes are censored at the time of death. The right-censoring rate is 54.5%. For each patient, the observed survival time and county of residence at diagnosis are recorded. The considered individual-level covariates include age at diagnosis and the stage: local, regional, or distant, where two dummy variables are created for regional and distant respectively, and the reference group is local. Zhou et al. (2015a) point out that some county-level socioeconomic factors (e.g., median household income, poverty level, education, rurality) are also potentially associated with breast cancer and argue that rural counties present more heterogeneity in access to quality care and screening for breast cancer. Therefore, we also include a county-level covariate "Rural-Urban Continuum Codes" (RUCC) measuring degree of urbanization; see Zhou et al. (2015a) for a detailed description.

We fit each of the PH, AFT and PO frailty models with a mixture of Polya trees prior on baseline survival $S_0(t)$ and the ICAR prior on the frailties $\gamma \sim \text{ICAR}(\lambda)$, where the PH is centered at the Weibull $G_\theta(t) = 1 - \exp\left\{-(e^{\theta_1}t)^{\exp(\theta_2)}\right\}$ and the AFT and PO are centered at the log-logistic $G_\theta(t) = 1 - \{1 + (e^{\theta_1}t)^{\exp(\theta_2)}\}^{-1}$. We consider the following prior settings: $J = 4$, $c \sim \Gamma(5,1)$, $\theta \sim N_2(\hat{\theta}, \hat{\mathbf{V}})$, $\beta \sim N_p(\hat{\beta}, 30\hat{\Sigma})$ and $\lambda \sim \Gamma(1,1)$, where $\hat{\theta}$, $\hat{\beta}$, $\hat{\mathbf{V}}$ and $\hat{\Sigma}$ are maximum likelihood estimates from the underlying parametric model. Using the same priors, we also fit the above models with Gaussian exchangeable frailties and without frailties. For all models considered, a burn-in of $100,000$ iterations is followed by a run of $100,000$ thinned down to $10,000$ iterations. All these models are fitted using the `survregbayes` function available in the package `spBayesSurv`.

Table 1 SEER breast cancer data: Posterior medians (95% credible intervals) of fixed effects from various models. Note the AFT model is parameterized as $S_{\mathbf{x}}(t) = S_0(e^{\mathbf{x}'\beta}t)$.

Model	Centered age	Regional stage	Distant stage	RUCC
PH/CAR	0.019 (0.013, 0.025)	0.26 (0.03, 0.48)	1.69 (1.45, 1.93)	-0.069 (-0.136, 0.002)
AFT/CAR	0.018 (0.011, 0.023)	0.22 (0.01, 0.43)	1.51 (1.26, 1.75)	-0.045 (-0.105, 0.013)
PO/CAR	0.030 (0.021, 0.038)	0.40 (0.12, 0.69)	2.59 (2.25, 2.95)	-0.087 (-0.174, -0.001)
PH/LDTFP	0.019 (0.013, 0.025)	0.27 (0.03, 0.49)	1.64 (1.43, 1.88)	-0.105 (-0.185, -0.041)

The LPML values under ICAR frailty PH, AFT and PO are -2226, -2228 and -2210, respectively, while the corresponding LPMLs are -2230, -2224 and -2214 under exchangeable frailty models and are -2230, -2228 and -2214 under

non-frailty models. We can observe that the ICAR frailty model has the best predictive ability within the context of either PH or PO, and the exchangeable frailty model performs best in terms of LPML under the AFT. Table 1 presents posterior means and equal-tailed 95% credible intervals (CI) for covariate effects under each of above model with ICAR frailties. All individual covariate effects are significant in each model. Higher age at diagnosis increases the hazard; e.g. a twenty-year increase in age is associated with an $\exp(0.019 \times 20) \approx 1.46$-fold increase in hazard. Using women with local stage of disease as the reference, the hazard rate of women of the same age who live in the same county will be $\exp(0.26) \approx 1.30$ times larger if their cancer is detected at the regional stage, and $\exp(1.69) \approx 5.42$ times larger if detected at the distant stage. Under the AFT assumption, among patients living in the same county and having same age, a woman with local stage typically survives $\exp(0.22) \approx 1.25$ times longer than a woman with regional stage, and $\exp(1.51) \approx 4.53$ times longer than a woman with distant stage. Finally, for the PO model, after adjusting for the age at diagnosis and the RUCC, the odds of dying from breast cancer before any time t are $\exp(0.40) \approx 1.49$ greater for regional stage versus local stage, and are $\exp(2.59) \approx 13.33$ greater for distant stage versus local stage. These findings are confirmed by Figure 1, which shows the fitted survival functions for women aged at 68.8 years and living a county with RUCC at 5 for distant and local stages under the three competing models and assuming a spatial frailty of zero. Turning to the county-level RUCC effect, only the PO model provides a significant result at the 0.05 level; living in more urban counties is associated with poorer survival after a breast cancer diagnosis on average.

Zhou et al. (2015a) fitted a PH model with LDTFP frailty terms using the package `DPpackage` and found more variability for frailties of rural counties. The resulting LPML is -2222 when RUCC is included into both the linear predictor and frailty terms. The pseudo Bayes factor for the LDTFP frailty model versus the ICAR frailty PH model is $\exp(2226-2222) \approx 55$, implying that the allowing frailties depending on RUCC improves the model's predictive ability about 55 times. Table 1 also shows the covariate effects under the LDTFP frailty PH model. An interesting finding is that now the RUCC effect becomes significant at the 0.05 level. This may be due to the fact that frailty distributions are covariate-dependent as shown in Figure 1(b). After controlling for individual covariates and county, the hazard rate of women living in urban counties (with RUCC $= 2$) will be $\exp(0.105 \times 7) \approx 2$ times larger than that of women in rural counties (with RUCC $= 9$).

5.2 Leukemia data

We consider a dataset on the survival of acute myeloid leukemia in $n = 1,043$ patients, analyzed by Henderson et al. (2002) fitting a multivariate gamma frailty PH model. This dataset is available for access in Fahrmeir and Kneib (2011). 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

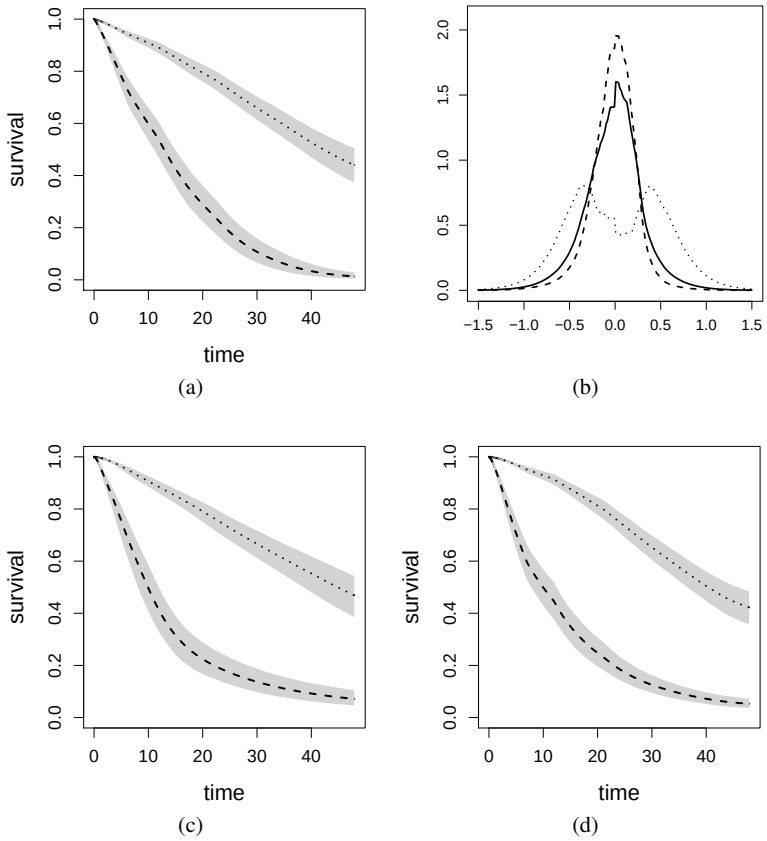

Fig. 1 SEER breast cancer data. Panels (a), (c) and (d) show estimated survival curves for women aged at 68.8 years and living a county with RUCC at 5 for distant (dashed lines) and local (dotted lines) stages, under PH, AFT and PO, respectively. The pointwise 95% credible bands are also displayed as grey areas. Panel (b) displays frailty densities for RUCC=2, 5 and 9, which are displayed as dashed, continuous and dotted lines, respectively.

(WBC) at diagnosis, and the Townsend score, for which higher values indicates less affluent areas. The censoring rate is 16%. Both exact residential locations of all patients and their administrative districts (24 districts that make up the whole region) are available. Therefore, we can fit both geostatistical and lattice models.

For the geostatistical case, we fit the copula model (19) proposed by Zhou et al. (2015b) using the function `spCopulaDDP`, where the marginal model $F_{\mathbf{x}}(\cdot)$ is defined via the LDDPM in (12) and the copula function C is specified through the Gaussian spatial copula in (20) assuming the exponential correlation function. We then use the function `spCopulaCoxph` to fit the copula model assuming a piecewise exponential PH model for $F_{\mathbf{x}}(\cdot)$, where the partition is based on $J = 20$ cutpoints with each a_k defined as the $\frac{k}{J}$th quantile the empirical distribution of observed

survival times (see Section 2.1). For comparison, standard non-spatial LDDPM and piecewise exponential PH models are also fitted using the functions `anovaDDP` and `indeptCoxph`, respectively. The default priors are considered for above models as suggested in Zhou et al. (2015b). Regarding the lattice case, we fit each of the Polya trees PH, AFT and PO models with ICAR frailties as in Section 5.1 using the function `survregbayes` and their corresponding non-frailty models, where the Polya trees are truncated at level $J = 5$. Finally, we fit the generalized AFT model (13) with and without ICAR frailties using the function `frailtyGAFT`, where ε_i is allowed to depends on age and WBC. We refer readers to Zhou et al. (2015c) for discussion of prior specifications and posterior samplings. For all models, we retain $10,000$ scans thinned from $50,000$ after a burn-in period of $10,000$ iterations.

The LPML measures for the copula with LDDPM, copula with piecewise exponential PH, PH, AFT and PO with Polya trees baselines and ICAR frailties, and generalized AFT with ICAR frailties are -5932, -5939, -5930, -5953, -5925 and -5936, respectively. Without spatial components, the above LPML values become -5934, -5941, -5934, -5950, -5925 and -5942. The PO models significantly outperform others from a predictive point of view regardless of whether spatial dependence is taken into account. Within the context of LDDPM and PH, the use of the Gaussian spatial copula slightly improves the model's predictive ability, indicating that the spatial dependence is relatively weak in this dataset. Under the framework of PH, the Polya trees prior works much better than piecewise exponential prior for modeling baseline functions. The AFT models provide the worst LPML values, while allowing the baseline varying with covariates (i.e., generalized AFT) can significantly improve the models' predictive ability; the Bayes factors for age and WBC effects on the baseline survival are 124 and 23 respectively under the ICAR frailty model, and are 73 and 31 under the non-frailty model.

For the copula LDDPM model, the posterior median of the nugget effect parameter θ_1 is 0.051 with the 95% CI $(0.000, 0.176)$, indicating that only 5% of the heterogeneity variance is explained by spatial effect on average. The posterior median of θ_2 is 0.831 with the 95% CI $(0.001,\ 3.075)$ indicates that the correlation decays by $1 - e^{-0.831} \approx 56\%$ for every kilometer increase in distance on average. However, given such a small value θ_1, the spatial decay becomes less important. Figure 2(a) shows the survival curves under the PO ICAR frailty model for female patients aged at 49 (25%th quantile) and aged at 74 (75%th quantile) holding other covariates at population averages, where we see that higher age is associated with lower survival probability. Figure 2(b) shows the baseline survival curves under the generalized AFT ICAR frailty model for female patients aged at 49 and aged at 74 holding WBC at its population average, where we can see that the baseline varies with age which clearly violates the AFT assumption.

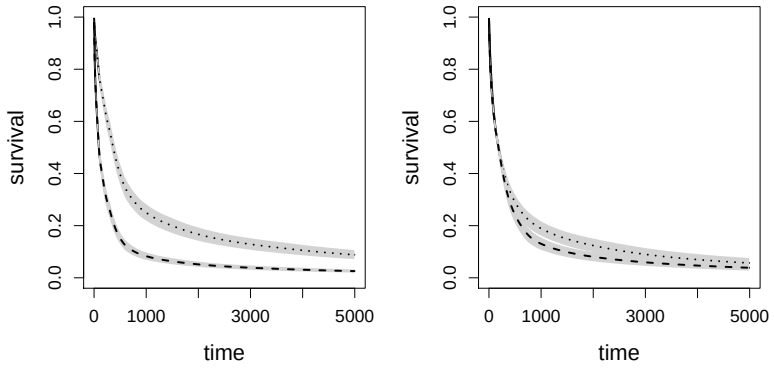

Fig. 2 Leukemia data. Panels (a) shows estimated survival curves for women aged at 49 years (dotted lines) and aged at 75 years (dashed lines), holding other covariates at population averages and frailties at zeros, under the PO model with ICAR frailties. Panels (b) shows estimated baseline survival curves for women aged at 49 years (dotted lines) and aged at 75 years (dashed lines), holding WBC at its population average and frailties at zeros, under the generalized AFT with ICAR frailties. The pointwise 90% credible bands are also displayed as grey areas.

6 Concluding remarks

We have reviewed commonly-used priors on baseline functions, semiparametric and nonparametric Bayesian survival models, and recent approaches for accommodating spatial dependence, both frailty and copula. Many R packages are discussed for implementation including `DPpackage`, `spBayesSurv`, `R2BayesX` and `spatsurv`. Two interesting data sets are illustrated, where both analyses show that PO models perform significantly better than all other models we considered including the PH, AFT and two generalizations of AFT.

Acknowledgements This work was supported by federal grants 1R03CA165110 and 1R03CA176739-01A1.

References

Aalen, O. O. (1980). A model for nonparametric regression analysis of counting processes. In *Mathematical Statistics and Probability Theory*, volume 2, pages 1–25. Springer-Verlag.

Aalen, O. O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine*, 8(8):907–925.

Alzola, C. and Harrell, F. (2006). *An Introduction to S and the Hmisc and Design Libraries*. Online manuscript available at http://biostat.mc.vanderbilt.edu/wiki/pub/Main/RS/sintro.pdf.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, 10(4):1100–1120.

Banerjee, S. and Carlin, B. P. (2003). Semiparametric spatio-temporal frailty modeling. *Environmetrics*, 14(5):523–535.

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. and Dey, D. K. (2005). Semiparametric proportional odds models for spatially correlated survival data. *Lifetime Data Analysis*, 11(2):175–191.

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.

Bárdossy, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research*, 42(11):1–12.

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.

Berger, J. O. and Guglielmi, A. (2001). Bayesian and conditional frequentist testing of a parametric model versus nonparametric alternatives. *Journal of the American Statistical Association*, 96(453):174–184.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, 66(3):429–436.

Burridge, J. (1981). Empirical Bayes analysis of survival time data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 43(1):65–75.

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.

Cai, B. and Meyer, R. (2011). Bayesian semiparametric modeling of survival data based on mixtures of B-spline distributions. *Computational Statistics & Data Analysis*, 55(3):1260–1272.

Carlin, B. P. and Hodges, J. S. (1999). Hierarchical proportional hazards regression models for highly stratified data. *Biometrics*, 55(4):1162–1170.

Chang, I.-S., Hsiung, C. A., Wu, Y.-J., and Yang, C.-C. (2005). Bayesian survival analysis using Bernstein polynomials. *Scandinavian Journal of Statistics*, 32(3):447–466.

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

Chen, Y. Q. and Jewell, N. P. (2001). On a general class of semiparametric hazards regression models. *Biometrika*, 88(3):687–702.

Chen, Y. Q. and Wang, M.-C. (2000). Analysis of accelerated hazards models. *Journal of the American Statistical Association*, 95(450):608–618.

Cheng, S. C., Wei, L. J., and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, 82(4):835–845.

Chernoukhov, A. (2013). Bayesian Spatial Additive Hazard Model. *Electronic Theses and Dissertations*. Paper 4965.

Christensen, R. and Johnson, W. (1988). Modeling accelerated failure time with a Dirichlet process. *Biometrika*, 75(4):693–704.

Clayton, D. (1991). A monte carlo method for Bayesian inference in frailty models. *Biometrics*, 47(2):467–485.

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. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.

Dabrowska, D. M. and Doksum, K. A. (1988). Estimation and testing in a two-sample generalized odds-rate model. *Journal of the American Statistical Association*, 83(403):744–749.

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

Dasgupta, P., Cramb, S. M., Aitken, J. F., Turrell, G., and Baade, P. D. (2014). Comparing multilevel and Bayesian spatial random effects survival models to assess geographical inequalities in colorectal cancer survival: a case study. *International Journal of Health Geographics*, 13(1):36.

de Boor, C. (2001). *A Practical Guide to Splines*. Applied Mathematical Sciences, Vol. 27. Springer-Verlag: New York.

De Iorio, M., Johnson, W. O., Müller, P., and Rosner, G. L. (2009). Bayesian nonparametric nonproportional hazards survival modeling. *Biometrics*, 65(3):762–771.

Devarajan, K. and Ebrahimi, N. (2011). A semi-parametric generalization of the Cox proportional hazards regression model: Inference and applications. *Computational Statistics & Data Analysis*, 55(1):667–676.

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.

Dukić, V. and Dignam, J. (2007). Bayesian hierarchical multiresolution hazard model for the study of time-dependent failure patterns in early stage breast cancer. *Bayesian Analysis*, 2:591–610.

Dunson, D. B. and Herring, A. H. (2005). Bayesian model selection and averaging in additive and proportional hazards. *Lifetime Data Analysis*, 11:213–232.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–102.

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

Fahrmeir, L. and Kneib, T. (2011). *Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data*. Oxford University Press.

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

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics*, 2:615–629.

Finley, A. O., Sang, H., Banerjee, S., and Gelfand, A. E. (2009). Improving the performance of predictive process modeling for large datasets. *Computational statistics & data analysis*, 53(8):2873–2884.

Furrer, R., Genton, M. G., and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523.

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

Gelfand, A. E. and Mallick, B. K. (1995). Bayesian analysis of proportional hazards models built from monotone functions. *Biometrics*, 51:843–852.

Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*, 87:942–951.

Griffin, J. (2010). Default priors for density estimation with mixture models. *Bayesian Analysis*, 5:45–64.

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. and Johnson, W. O. (2002). Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association*, 97(460):1020–1033.

Hanson, T., Kottas, A., and Branscum, A. (2008). Modelling stochastic order in the analysis of receiver operating characteristic data: Bayesian nonparametric approaches. *Journal of the Royal Statistical Society: Series C*, 57:207–225.

Hanson, T. E. (2006a). Inference for mixtures of finite Polya tree models. *Journal of the American Statistical Association*, 101(476):1548–1565.

Hanson, T. E. (2006b). Modeling censored lifetime data using a mixture of gammas baseline. *Bayesian Analysis*, 1:575–594.

Hanson, T. E., Branscum, A., and Johnson, W. O. (2005). Bayesian nonparametric modeling and data analysis: An introduction. In Dey, D. and Rao, C., editors, *Bayesian Thinking: Modeling and Computation (Handbook of Statistics, volume 25)*, pages 245–278. Elsevier: Amsterdam.

Hanson, T. E., Branscum, A., and Johnson, W. O. (2011). Predictive comparison of joint longitudinal–survival modeling: a case study illustrating competing approaches. *Lifetime Data Analysis*, 17:3–28.

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.

Hanson, T. E. and Yang, M. (2007). Bayesian semiparametric proportional odds models. *Biometrics*, 63(1):88–95.

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.

Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, 18:1259–1294.

Hutton, J. L. and Monaghan, P. F. (2002). Choice of parametric accelerated life and proportional hazards models for survival data: Asymptotic results. *Lifetime Data Analysis*, 8:375–393.

Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer-Verlag.

Jara, A., Hanson, T., Quintana, F., Müller, P., and Rosner, G. (2011). DPpackage: Bayesian semi- and nonparametric modeling in R. *Journal of Statistical Software*, 40(5):1.

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

Jara, A., Lesaffre, E., De Iorio, M., and Quitana, F. (2010). Bayesian semiparametric inference for multivariate doubly-interval-censored data. *The Annals of Applied Statistics*, 4(4):2126–2149.

Johnson, W. O. and Christensen, R. (1989). Nonparametric bayesian analysis of the accelerated failure time model. *Statistics and Probability Letters*, 8:179–184.

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

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.

Kay, R. and Kinnersley, N. (2002). On the use of the accelerated failure time model as an alternative to the proportional hazards model in the treatment of time to event data: A case study in influenza. *Drug Information Journal*, 36:571–579.

Kneib, T. (2006). *Mixed model based inference in structured additive regression*. Ludwig-Maximilians-Universität München.

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

Koenker, R. (2008). Censored quantile regression redux. *Journal of Statistical Software*, 27(6):1–25.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15:143–156.

Komárek, A. and Lesaffre, E. (2007). Bayesian accelerated failure time model for correlated censored data with a normal mixture as an error distribution. *Statistica Sinica*, 17:549–569.

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*, 103:523–533.

Kottas, A. and Gelfand, A. E. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, 95:1458–1468.

Kuo, L. and Mallick, B. (1997). Bayesian semiparametric inference for the accelerated failure-time model. *Canadian Journal of Statistics*, 25:457–472.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13:183–212.

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

Lavine, M. (1994). More aspects of Polya tree distributions for statistical modelling. *The Annals of Statistics*, 22:1161–1176.

Li, J. (2010). Application of copulas as a new geostatistical tool. *Dissertation*.

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*, in press.

Li, L., Hanson, T., and Zhang, J. (2015b). Spatial extended hazard model with application to prostate cancer survival. *Biometrics*, in press.

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*, in press.

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. (2012). *Bayesian analysis of spatial and survival models with applications of computation techniques*. PhD thesis, University of Missouri–Columbia.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *Annals of Statistics*, 12:351–357.

Martinussen, T. and Scheike, T. H. (2006). *Dynamic Regression Models for Survival Data*. Springer-Verlag.

McKinley, T. J. (2007). *Spatial survival analysis of infectious animal diseases*. PhD thesis, University of Exeter.

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

Müller, P. and Quintana, F. A. (2004). Nonparametric Bayesian data analysis. *Statistical Science*, 19(1):95–110.

Murphy, S. A., Rossini, A. J., and van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*, 92:968–976.

Neal, R. M. (2000). Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9:249–265.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, 2nd edition.

Nieto-Barajas, L. E. (2013). Lévy-driven processes in Bayesian nonparametric inference. *Boletín de la Sociedad Matemática Mexicana*, 19:267–279.

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

Orbe, J., Ferreira, E., and Núñez Antón, V. (2002). Comparing proportional hazards and accelerated failure time models for survival analysis. *Statistics in Medicine*, 21(22):3493–3510.

Pan, C., Cai, B., Wang, L., and Lin, X. (2014). Bayesian semi-parametric model for spatial interval-censored survival data. *Computational Statistics & Data Analysis*, 74:198–209.

Petrone, S. (1999a). Bayesian density estimation using Bernstein polynomials. *The Canadian Journal of Statistics*, 27:105–126.

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

Reid, N. (1994). A conversation with Sir David Cox. *Statistical Science*, 9:439–455.

Ryan, T. and Woodall, W. (2005). The most-cited statistical papers. *Journal of Applied Statistics*, 32(5):461–474.

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.

Scharfstein, D. O., Tsiatis, A. A., and Gilbert, P. B. (1998). Semiparametric efficient estimation in the generalized odds-rate class of regression models for right-censored time-to-event data. *Lifetime data analysis*, 4(4):355–391.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650.

Sharef, E., Strawderman, R. L., Ruppert, D., Cowen, M., and Halasyamani, L. (2010). Bayesian adaptive B-spline estimation in proportional hazards frailty models. *Electronic Journal of Statistics*, 4:606–642.

Sinha, D. and Dey, D. K. (1997). Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association*, 92:1195–1212.

Sinha, D., McHenry, M. B., Lipsitz, S. R., and Ghosh, M. (2009). Empirical bayes estimation for additive hazards regression models. *Biometrika*, 96(3):545–558.

Smith, M. S. (2013). Bayesian approaches to copula modelling. *Bayesian Theory and Applications*, pages 336–358.

Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *Journal of the American Statistical Association*, 71:897–902.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag Inc.

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.

Walker, S. G. and Mallick, B. K. (1997). Hierarchical generalized linear models and frailty models with Bayesian nonparametric mixing. *Journal of the Royal Statistical Society, Series B: Methodological*, 59:845–860.

Walker, S. G. and Mallick, B. K. (1999). A Bayesian semiparametric accelerated failure time model. *Biometrics*, 55(2):477–483.

Wang, L. and Dunson, D. B. (2011). Semiparametric Bayes' proportional odds models for current status data with underreporting. *Biometrics*, 67(3):1111–1118.

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*.

Yang, S. (1999). Censored median regression using weighted empirical survival and hazard functions. *Journal of the American Statistical Association*, 94:137–145.

Yang, S. and Prentice, R. L. (1999). Semiparametric inference in the proportional odds regression model. *Journal of the American Statistical Association*, 94:125–136.

Yin, G. and Ibrahim, J. G. (2005). A class of Bayesian shared gamma frailty models with multivariate failure time data. *Biometrics*, 61:208–216.

Ying, Z., Jung, S. H., and Wei, L. J. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association*, 90:178–184.

Zellner, A. (1983). Applications of Bayesian analysis in econometrics. *The Statistician*, 32:23–34.

Zhang, J. and Lawson, A. B. (2011). Bayesian parametric accelerated failure time spatial model and its application to prostate cancer. *Journal of Applied Statistics*, 38(3):591–603.

Zhang, J., Peng, Y., and Zhao, O. (2011). A new semiparametric estimation method for accelerated hazard model. *Biometrics*, 67:1352–1360.

Zhang, M. and Davidian, M. (2008). "Smooth" semiparametric regression analysis for arbitrarily censored time-to-event data. *Biometrics*, 64(2):567–576.

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., Hanson, T., Jara, A., and Zhang, J. (2015a). Modelling county level breast cancer survival data using a covariate-adjusted frailty proportional hazards model. *The Annals of Applied Statistics*, in press.

Zhou, H., Hanson, T., and Knapp, R. (2015b). Marginal Bayesian nonparametric model for time to disease arrival of threatened amphibian populations. *Biometrics*, in press.

Zhou, H., Hanson, T., and Zhang, J. (2015c). Generalized accelerated failure time spatial frailty model for arbitrarily censored data. *Lifetime Data Analysis*, in revision.