Published in final edited form as:

Biometrics. 2011 June ; 67(2): 391–403. doi:10.1111/j.1541-0420.2010.01468.x.

Spatially Dependent Polya Tree Modeling for Survival Data

Luping Zhao and

Eli Lilly and Company, Indianapolis, Indiana, U.S.A

Timothy E. Hanson

Department of Statistics, University of South Carolina, Columbia, South Carolina, U.S.A

Luping Zhao: zhao0117@umn.edu; Timothy E. Hanson: hans0058@umn.edu

Summary

With the proliferation of spatially oriented time-to-event data, spatial modeling in the survival context has received increased recent attention. A traditional way to capture a spatial pattern is to introduce frailty terms in the linear predictor of a semiparametric model, such as proportional hazards or accelerated failure time. We propose a new methodology to capture the spatial pattern by assuming a prior based on a mixture of spatially dependent Polya trees for the baseline survival in the proportional hazards model. Thanks to modern Markov chain Monte Carlo (MCMC) methods, this approach remains computationally feasible in a fully hierarchical Bayesian framework. We compare the spatially dependent mixture of Polya trees (MPT) approach to the traditional spatial frailty approach, and illustrate the usefulness of this method with an analysis of Iowan breast cancer survival data from the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute. Our method provides better goodness of fit over the traditional alternatives as measured by log pseudo marginal likelihood (LPML), the deviance information criterion (DIC) and full sample score (FSS) statistics.

Keywords

Breast cancer; Conditionally autoregressive (CAR) model; Log pseudo marginal likelihood (LPML); Mixture of Polya trees; Nonparametric modeling; Proportional hazards

1. Introduction

In biomedical and epidemiological studies, time-to-event data often contain geographical or spatial information, such as the county or ZIP code of residence. With the proliferation of spatially oriented survival data sets, spatial modeling has received recent increased attention. If the interest of investigation is only on population-level effects, not on the spatial pattern, a stratified analysis in SAS Proc PHREG can be employed. In this way, the spatial-level differences in survival are treated as nuisances. When interest resides in how survival changes across space as well as estimating population-level risk factors, an appropriate estimate of the spatial pattern is essential.

Hierarchical exchangeable frailty models were introduced by Vaupel, Manton, and Stallard (1979). These models, and variants including partial exchangeability and other (e.g. spatial) structures are highly flexible and hundreds of papers have been published using this basic

Supplementary Materials

Web Appendix A, referenced in Section 2.4, is available under the Paper Information link at the Biometrics website <http://www.biometrics.tibs.org>.

idea to accommodate strata-specific heterogeneity. As traditionally implemented, frailty terms are introduced into the linear predictor in models such as the accelerated failure time (AFT) model (e.g. Komárek, Lesaffre, and Legrand, 2007; Hanson et al., 2003), the proportional odds (PO) model (e.g. Banerjee and Dey, 2005; Zeng, Lin, and Yin, 2005), or the proportional hazards model (e.g. Banerjee and Carlin, 2003; also see Ibrahim, Chen, and Sinha, 2001, Chapter 4). Conditional on the frailty term, the models retain their interpretation in terms of constants of proportionality. For example, the PH model conditional on frailty γ_i assumes the survival curve for individual j in strata i , with covariates

$\mathbf{z}_{ij}$, follows $S_{ij}(t) = S_0(t)^{\exp(\mathbf{z}_{ij}'\beta + \gamma_i)}$ where $S_0(t)$ is a baseline survival curve. Note this implies

$S_{ij}(t) = S_{0i}(t)^{\exp(\mathbf{z}_{ij}'\beta)}$ where $S_{0i}(t) = S_0(t)^{\exp(\gamma_i)}$. Spatial frailty models can follow either geostatistical or lattice frameworks depending on how the geographical regions are indexed. In the geostatistical case, the regions are indexed to specific geographical coordinates, and the frailty vector $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)'$ is typically assumed to follow a normal distribution with mean $\mathbf{0}$ and a non-diagonal covariance matrix with entries depending on distances between the geographical locations. Lattice models define the dependence structure through the adjacency of the geographic regions.

Many authors have motivated and argued for spatially varying frailties in the conditional PH model, including Henderson, Shimakura, and Gorst (2002), Banerjee, Wall, and Carlin (2003), Li and Ryan (2004), Bastos and Gamerman (2006), Hennerfeind, Brezger, and Fahrmeir (2006), and Zhao, Hanson, and Carlin (2009), among others. Banerjee, Carlin, and Gelfand (2004) devote an entire chapter to spatial frailty modeling in survival, for the same reasons that random effects models are used: often information is available on cluster membership that may affect the outcome of interest beyond the recorded covariates; this membership is effectively blocked on. In survival modeling, Banerjee et al. list "...access to care, willingness of the population to seek care...overall quality of available health or hospital care..." and other event-specific underlying factors as potentially being more similar in proximal regions. Similarly, Henderson et al. (2002) list "...patient management differences between treatment centers...or due to background variation in population or environmental characteristics, necessitating further epidemiologic study..." as proximal effects. Thus, the resulting smoothed maps could be used to generate hypotheses concerning district-level access and quality of care, travel time to clinics, et cetera. Although spatial frailties serve as proxies to unmeasured regional-level covariates, they are less precise adjustments than, say, random effects accounting for repeated measures on the same individual, since regional-level covariates (such as shortest distance to a clinic) are unlikely to sharply change at areal boundaries.

In the traditional conditional frailty approach, such as in Banerjee and Carlin (2003), the PH models include simple frailties with one overall baseline distribution. Since spatial frailties are modeled as an additive term in the linear predictor, survival model assumptions are confounded with the frailty terms. One approach to relax and test such a model is to remove the frailty terms from the linear predictor and instead allow each geographic subregion to have a baseline distribution correlated to that of nearby regions. We consider a smoothly varying baseline $S_{0i}(t)$ that flexibly captures spatial survival trends, and allows for testing

the adequacy of simpler frailty models. The model, $S_{ij}(t) = S_{0i}(t)^{\exp(\mathbf{z}_{ij}'\beta)}$, retains the same conditional interpretation as the frailty models, but assumes considerably less structure. The spatially varying, or *dependent* baselines $\{S_{01}(\cdot), \dots, S_{0n}(\cdot)\}$ are centered at an overall parametric log-logistic family with parameters $\boldsymbol{\theta}$, but nonparametrically spatially "refined" via conditional probabilities that follow a spatial-lattice smoothing model. For each subregion i , the baseline survival function $S_{0i}(\cdot)$ approximately follows a MPT prior with Polya tree probabilities $\mathcal{X}_i$. Dependence is induced through proper, independent conditional

autoregressive (CAR, Besag, 1974) priors on the logit transformed Polya tree conditional probabilities.

The Polya tree probabilities π_{i_1} and π_{i_2} defining baseline survival functions $S_{0i_1}(\cdot)$ and $S_{0i_2}(\cdot)$ for adjacent counties i_1 and i_2 are stochastically smoothed to be similar; probabilities from counties far away from each other may not be. This specification marginally specifies a finite tailfree process for each county, centered at an overall log-logistic baseline. This model allows for smoothed areal (county-level) baseline survival distributions within the context of a population-level survival model. The spatial terms do not enter into the linear predictor, but the model retains a simple interpretation conditioned on the county.

As implemented, existing approaches to dependent processes have focused on regression with a spatially varying component; either spatially referenced data $y(\mathbf{s})$ or, in the case of a generalized linear model with link $g(\cdot)$, $g^{-1}[E\{y(\mathbf{s})\}]$ follows the process $\mu(\mathbf{s}) + \mathbf{x}(\mathbf{s})'\boldsymbol{\beta} + e(\mathbf{s})$, where $\mu(\mathbf{s})$ is some development of a dependent stick-breaking prior, the $\mathbf{x}(\mathbf{s})$ are covariates, and the $e(\mathbf{s})$ are *iid* white noise, typically $N(0, \sigma^2)$; see e.g. expressions (1) in Kottas, Duan, and Gelfand (2008), (6) in Griffin and Steel (2006), (1) in Dunson, Pillai, and Park (2007), and (4) in Reich and Fuentes (2007). The inclusion of $e(\mathbf{s})$ is actually *required* to use modifications of the now well-developed computational theory available for fitting Dirichlet process mixture (DPM) models. This essentially amounts to requiring a nugget effect in the spatial or covariate process, but practically translates into a process that behaves similarly to a finite mixture model with component variances fixed at σ^2 and locations or weights varying smoothly in space or time. This built-in linear structure implies that dependent stick-breaking priors lend themselves naturally to accelerated failure time (AFT) models, but not necessarily to proportional hazards (PH) which relate survival to covariates in a nonlinear way (Hanson, 2006).

We compared our proposed spatially dependent Polya tree PH model to the corresponding traditional PH spatial frailty model using an analysis of a cohort of 5786 women diagnosed with malignant breast cancer starting in 1989 ending 1991, with follow-up continued through the end of 2003. The log pseudo marginal likelihood (LPML), deviance information criterion (DIC) and full sample score (FSS) statistics suggest that the spatially dependent Polya tree baseline PH model provides better goodness of fit to the data and is predictively superior to the corresponding traditional PH spatial frailty model with a common survival baseline. The models described herein allow consideration of how survival changes geographically; e.g. we examine spatial patterns of survival 10th percentiles from subregion-level choropleth maps.

The remainder of our paper is organized as follows. Section 2 gives a detailed description of the proposed statistical models, including computational details related to MCMC implementation of the mixture of spatially dependent Polya trees. Section 3 then offers a detailed analysis of the SEER dataset, including model comparison, parameter estimation, mapping of spatially smoothed, county-specific baseline hazard curves for some specific counties. Section 4 provides an analysis of simulated data. Finally, Section 5 discusses our findings, and offers directions for future work in this area.

2. Statistical models

In the survival literature, the traditional method for accounting for spatial dependence is to introduce spatial frailty terms in the linear predictor. Zhao, Hanson, and Carlin (2009) fit traditional PH spatial frailty models with a common MPT prior for the survival baseline and show both the spatial frailty terms and the nonparametric MPT prior improve model fit. In this paper, we wish to explore a more general approach that captures the spatial pattern through a spatially correlated survival baseline across subregions. We assume for each

subregion i , the baseline survival function S_{0i} approximately follows a Polya tree prior centered at the log-logistic parametric family, and these baseline survival functions are spatially correlated. That is, each subregion has a spatially dependent baseline which borrows information from adjacent subregions, but is only slightly affected by subregions far away. We compare the proposed model with the simpler PH spatial frailty models of Hennerfeind et al. (2006) and Zhao, Hanson, and Carlin (2009).

2.1 Mixture of Polya trees prior

We consider a PH model with spatially dependent MPT priors on subregion (i.e., county) level survival baselines. Lavine (1992) gives a general definition of a finite Polya tree prior. We consider a particular prior suitable for the spatial model; variants of this prior have been used in the literature (Lavine 1992, 1994; Walker and Mallick, 1997, 1999; Hanson and Johnson, 2002; Zhao, Hanson, and Carlin, 2009).

Let $G_{\theta}(\cdot)$ be the cumulative distribution function from an absolutely continuous parametric family with support on the positive reals $\mathbb{R}^+$, indexed by θ . A Polya tree prior on a random distribution G is constructed from a set of increasingly refined partitions of $\mathbb{R}^+$. At each level $l \in \{1, \dots, M\}$ partition $\mathbb{R}^+$ into 2^l sets $B_{\theta}(l, k) = (G_{\theta}^{-1}((k-1)2^{-l}), G_{\theta}^{-1}(k2^{-l}))$, where $k = 1, \dots, 2^l$. Note that $B_{\theta}(l, k) = B_{\theta}(l+1, 2k-1) \cup B_{\theta}(l+1, 2k)$. Let $X_{l,2k-1} = G\{B_{\theta}(l, 2k-1) | B_{\theta}(l-1, k)\}$ and $X_{l,2k} = 1 - X_{l,2k-1} = G\{B_{\theta}(l, 2k) | B_{\theta}(l-1, k)\}$ be the conditional probabilities of being in either of the two “offspring” sets of $B_{\theta}(l-1, k)$. Assume G follows G_{θ} on sets in the finest partition $\{B_{\theta}(M, k)\}$. The conditional probability vectors at MPT location $(l-1, k)$, $(X_{l,2k-1}, X_{l,2k})$, are independent Dirichlet:

$(X_{l,2k-1}, X_{l,2k}) \stackrel{\text{ind.}}{\sim} \text{Dirichlet}(c\rho(l), c(l))$, except $(X_{1,1}, X_{1,2}) = (0.5, 0.5)$ for identifiability. This implies $X_{l,2k-1} \sim \text{beta}(c\rho(l), c(l))$ and $X_{l,2k} = 1 - X_{l,2k-1}$. Given the M partitions

$\{B_{\theta}(l, k)\}_{k=1}^{2^l}$, $l = 1, \dots, M$, a particular finite Polya tree prior with parameters c , $\rho(\cdot)$, and G_{θ} is defined as in Table 1 for $l = 1, 2, 3$.

The parameter c is related to how quickly data “take over” the prior. As c tends to zero the posterior baseline is almost entirely data-driven; as c tends to infinity we obtain a fully parametric analysis. The function $\rho(\cdot)$ typically increases; $\rho(l) = l^2$ is often used. In this paper we center G at the log-logistic distribution, so $G_{\theta}(t) = 1 - \{1 + e^{\eta t \exp(\alpha)}\}^{-1}$ for $t > 0$.

This construction involves one overall weight parameter c . However, we can let c vary by level l , or even allow each pair of conditional probabilities to have their own $c_{l,k}$. This more closely mirrors the original definition in Lavine (1992) and allows for varying amounts of flexibility for G relative to G_{θ} across $\mathbb{R}^+$. For example, in heavily censored survival data, as we consider here, there is naturally more information for the shape of G toward smaller (uncensored) data points versus the upper tail.

Define the independent, random vectors of conditional probabilities $\mathcal{X} = \{X_{l,k}\}$, where $l = 1, \dots, M$ and $k = 1, \dots, 2^l$. For each level $l = 1, \dots, M$ and partition interval $k = 1, \dots, 2^l$,

$G\{B_{\theta}(l, k) | \mathcal{X}\} = \prod_{m=1}^l X_{m, \lceil k2^{m-l} \rceil}$, where $\lceil x \rceil$ is the ceiling function: the smallest integer greater than or equal to x . Assume that the random distribution G has the finite Polya tree prior described above with parameters c , $\rho(\cdot)$, and G_{θ} , written $G \sim PT(c, \rho, G_{\theta})$. Given $\mathcal{X}$ and θ , G is known. Define the vector of probabilities $\mathbf{p}_{\mathcal{X}} = (p_{\mathcal{X}}(1), p_{\mathcal{X}}(2), \dots, p_{\mathcal{X}}(2^M))'$ as

$$p_X(k) = G\{B_\theta(M, k) | \mathcal{X}\} = \prod_{l=1}^M X_{l, \lceil k2^{l-M} \rceil}. \quad (1)$$

Depending on the context, $G(\cdot)$ is used to denote either the measure of a set or the cdf at a point. The cdf $G(x | \mathcal{X}, \theta)$ is computed from (1) by

$$G(x | \mathcal{X}, \theta) = \left\{ \sum_{k=1}^{k_\theta(M, x)-1} p_X(k) \right\} + p_X(k_\theta(M, x)) \{2^M G_\theta(x) - k_\theta(M, x) + 1\}, \quad (2)$$

where $k_\theta(M, x) = \text{Int}\{2^M G_\theta(x) + 1\}$. The density associated with $G(x | \mathcal{X}, \theta)$ is given by

$$g(x | \mathcal{X}, \theta) = 2^M \sum_{k=1}^{2^M} p_X(k) g_\theta(x) I_{B_\theta(M, k)}(x) = 2^M p_X(k_\theta(M, x)) g_\theta(x). \quad (3)$$

The quantile function associated with G is given by $G^{-1}(q | \mathcal{X}, \theta) = G_\theta^{-1} \left\{ \frac{q - \sum_{k=1}^N p_X(k) + N p_X(N)}{2^M p_X(N)} \right\}$,

where N is such that $\sum_{k=1}^{N-1} p_X(k) < q \leq \sum_{k=1}^N p_X(k)$. This quantile function is used to find posterior estimates and credible intervals of survival percentiles for given covariates as in Section 3.2.

A mixture of finite Polya trees prior is induced by allowing θ to be random. The prior is written $G \sim \int PT(c, \rho, G_\theta) p(\theta) d\theta$. Letting θ be random smooths over partitioning effects inherent in a simple Polya tree (Lavine, 1992; Hanson, 2006) and centers G at a parametric family rather than a single fixed distribution.

2.2 Spatial smoothing of baseline survival

The spatially-smoothed baseline survival functions $\{S_{01}(\cdot), \dots, S_{0n}(\cdot)\}$ are built from $2^M - 2$ independent CAR priors on transformed conditional probabilities defining the Polya tree. Let $\mathcal{X}_i$ be the parameters defining $S_{0i}(\cdot)$. The conditional probability associated with $B_\theta(l, k)$ in strata i is denoted $X_{l,k}(i) \in \mathcal{X}_i$. Define $\mathbf{X}_{l,k} = (X_{l,k}(1), \dots, X_{l,k}(n))'$. These probabilities are logistic-transformed mean-zero Gaussian random variables $Y_{l,k}(i)$,

$$X_{l,2k-1}(i) = \frac{e^{Y_{l,2k-1}(i)}}{1 + e^{Y_{l,2k-1}(i)}} \text{ and } X_{l,2k}(i) = \frac{1}{1 + e^{Y_{l,2k-1}(i)}} \text{ where } k=1, \dots, 2^{l-1}. \quad (4)$$

The proper CAR distribution (e.g. Banerjee, Carlin, and Gelfand, 2004, p. 80) is defined in terms of conditional probabilities. For k odd let

$$Y_{l,k}(i) | \lambda, \{Y_{l,k}(j)\}_{j \neq i} \sim N \left(r \sum_{j=1}^n \frac{\omega_{ij}}{\omega_{i+}} Y_{l,k}(j), \frac{1}{\lambda \rho(l) \omega_{i+}} \right), i=1, \dots, n. \quad (5)$$

where $\omega_{ij} = 1$ if subregion j is adjacent to subregion i , $\omega_{ij} = 0$ if subregion j is not a neighbor of subregion i , and $\omega_{ii} = 0$ since subregion i is not a neighbor of itself. Define $\omega_{i+} = \sum_{j=1}^n \omega_{ij}$, the number of neighbors of subregion i ; λ is a precision parameter, and r is “propriety parameter” such that $r < 1$. By definition, the conditional probability of $Y_{l,k}(i)$ given all the other $Y_{l,k}(j)$ ’s depends only on its neighbors.

Let $\mathbf{W}$ be the symmetric adjacency matrix with $(\mathbf{W})_{ij} = \omega_{ij}$. Conditions on $\mathbf{W}$ necessary for the compatibility of conditional distributions are satisfied; by Brook’s Lemma the joint prior distribution is multivariate normal. Where $\mathbf{Y}_{l,k} = (Y_{l,k}(1), \dots, Y_{l,k}(n))'$, we have

$$Y_{l,k} | \lambda \sim N_n \left(\mathbf{0}, \frac{1}{\lambda \rho(l)} (\mathbf{D}_\omega - r \mathbf{W})^{-1} \right), \quad (6)$$

where $\mathbf{D}_\omega$ is diagonal with $(\mathbf{D}_\omega)_{ii} = \omega_{i+}$. Therefore $\mathbf{Y}_{l,k}$ has mean $\mathbf{0}$ and covariance matrix $\{\lambda \rho(l)\}^{-1} (\mathbf{D}_\omega - r \mathbf{W})^{-1}$.

Under the prior (6), $E\{Y_{l,k}(i)\} = 0$, so $E\{X_{l,k}(i)\} = 0.5$ and therefore G is centered at G_θ in that given θ , $E\{G(A)\} = G_\theta(A)$ for any measurable $A \subset \mathbb{R}^+$. Zhao and Hanson (2008) show that assuming $Y_{l,k}(i) \sim N(0, 2/c\rho(l))$ approximates $X_{l,k}(i) \sim \text{beta}(c\rho(l), c\rho(l))$. This is a marginal specification designed to mimic a “traditional” Polya tree prior as closely as possible through the use of a transformed process $\{Y_{l,k}(i) : i \in \mathcal{I}\}$. We attempt to do something similar through the conditional specification (5), but the marginal distribution is actually $Y_{l,k}(i) \sim N(0, (\lambda \rho(l))^{-1} \sigma_{ii})$ where σ_{ij} is the ij^{th} element of $(\mathbf{D}_\omega - r \mathbf{W})^{-1}$. Thus the conditional specification translates marginally to an approximate Polya tree prior $S_{0i} \sim PT(c_i, \rho, G_\theta)$ where $c_i = 2\lambda/\sigma_{ii}$, although a realization of the spatial process $S_{0i}(\cdot)$ is technically a member of the class of tailfree processes (Ferguson, 1974).

The value σ_{ii} can be bounded using results for diagonally dominant and/or positive definite symmetric matrices in terms of matrix elements or eigen values (e.g. Benzi and Golub, 1999; Theorem 3.1). However, for a particular incidence matrix $\mathbf{W}$, the exact diagonal elements can be computed for fixed values of r . The variance for strata with few neighbors will necessarily be larger than for strata with more neighbors. The induced precision $c_i = 2\lambda/\sigma_{ii}$ will vary across subregions. We are untroubled by this, in part because we assume λ has a vague prior to begin with: an equal-tailed 95% probability interval for $\lambda \sim \Gamma(0.1, 0.1)$ is about $(10^{-15}, 10)$. However, if one wishes to fix the marginal precision to be constant across subregions, $c_1 = \dots = c_n = c$, Besag and Kooperberg (1995) provide a roadmap. Additional flexibility is gained by allowing each conditional probability process $\{X_{l,2k-1}(i) : i \in \mathcal{I}\}$ to have a unique precision parameter $\lambda_{l,2k-1}$, $k = 1, \dots, 2^{l-1}$. This allows, for example, a more nonparametric or “bumpy” density shape near earlier survival times than near later times where more censoring occurs and there is relatively less data to estimate density shape. However, taking $\lambda_{l,2k-1}$ separate across $l = 2, \dots, M$ and $k = 1, \dots, 2^{l-1}$ precludes the simpler approximation $S_{0i} \sim PT(c_i, \rho, G_\theta)$ where $c_i = 2\lambda/\sigma_{ii}$ in favor of a more general Polya tree prior.

2.3 Hierarchical model

We assume PH conditional on strata i . Let $\mathbf{z}_{ij}$ be a p -dimensional vector of explanatory covariates associated with the j^{th} individual in group i , $j = 1, \dots, n_i$, $i = 1, \dots, n$, and let $S_{ij}(\cdot)$ be the associated survival function. We consider patients grouped at the county level, so that n is the number of counties. The most general model is

$$T_{ij}|\beta, S_{0i} \stackrel{ind.}{\sim} S_{ij}(\cdot) = S_{0i}(\cdot)^{\exp(\mathbf{z}'_{ij}\beta)}, \quad (7)$$

$$\begin{aligned} \text{logit}\{X_{l,k}(i)\} &= Y_{l,k}(i), \text{ where} \\ Y_{l,k}|\lambda_{l,k} &\sim \text{Proper CAR}(\lambda_{l,k}, r), \text{ i.e.} \end{aligned} \quad (8)$$

$$\begin{aligned} Y_{l,k}(i)|\lambda_{l,k}, \{Y_{l,k}(s)\}_{s \neq i} &\sim N\left(r \sum_{s=1}^n \frac{\omega_{is}}{\omega_{i+}} Y_{l,k}(s), \frac{1}{\lambda_{l,k}\omega_{i+}p(l)}\right), \\ \lambda_{l,k} &\stackrel{iid}{\sim} \Gamma(0.1, 0.1), \end{aligned} \quad (9)$$

$$(\theta', \beta') \sim N_{p+2}(\mathbf{m}, \mathbf{S}). \quad (10)$$

In data analyses, we assume a vague but proper hyperprior $\mathbf{m} = \mathbf{0}$ and $\mathbf{S} = 1000\mathbf{I}_{p+2}$, but prior information could be elicited and implemented. For example, see Greenland and Christensen (2001) for an approach in the stratified proportional hazards model.

To recap, the a simple Polya tree prior yields $S_0(\cdot) = 1 - G(\cdot)$ given by (2) using the probabilities $p_{\mathcal{X}}(\cdot)$, given by (1), on the finest partition through the conditional probabilities $\mathcal{X}$. Beyond this, we model each conditional probability in $\mathcal{X}$ as spatially varying using logistically-transformed CAR prior distributions, so the conditional probabilities, and hence baseline survival, are smoothed in space.

The parameters λ and r control the tradeoff between smoothing and allowing for essentially independent S_{0i} across the subregions (i.e., a stratified analysis) in the proper CAR prior. We tried different values for the propriety parameter: $r = 0, 0.5, 0.9, 0.99$. As r tends to 0, the $\{S_{01}, \dots, S_{0n}\}$ become independent. As r tends to 1, the joint prior is improper yielding an intrinsically autoregressive prior.

We assume a different precision parameter $\lambda_{l,k}$ at different MPT location, but also consider the simpler model, $\lambda_{l,k} \equiv \lambda$, yielding $S_{0i}(\cdot)|\mathcal{X}, \theta, \lambda$ approximately distributed as $PT(c_i, \rho, G_\theta)$ where $\lambda \sim \Gamma(0.1, 0.1)$. For $r = 0.99$, we have $1.4 < 2/\sigma_{ii} < 4.7$ for the Iowa data analyzed in Section 3.2. Then $\lambda \sim \Gamma(0.1, 0.1)$ implies reasonable induced precision priors for $c_i = 2\lambda/\sigma_{ii}$. We call the dependent MPT model with common λ and $r = 0.99$ Model 1A. Note that as $\lambda \rightarrow \infty$ a PH model with the same parametric log-logistic baseline $S_{0i}(t) = S_0(t)$ is obtained.

2.4 Computational notes

Given $(\beta, \theta, \mathcal{X}_i)$ and covariates $\mathbf{z}_{ij}$ the pdf of the survival time from the j^{th} subject in strata i is given by $p_{ij}(t|\beta, \theta, \mathcal{X}_i) = \exp(\mathbf{z}'_{ij}\beta) S_{i0}(t)^{\exp(\mathbf{z}'_{ij}\beta)-1} g_{i0}(t)$, evaluated through (2) and (3). The Polya tree parameters $p(\mathcal{X}_i|\{\lambda_{l,k}\})$ follow the product of $2^M - 2$ independent logistic normal densities, obtained marginally from (6). We place a normal prior on $(\beta', \theta)'$, and gamma priors on the precisions $\lambda_{l,k} \stackrel{iid}{\sim} \Gamma(0.1, 0.1)$. For common $\lambda_{l,k} \equiv \lambda$, $\lambda \sim \Gamma(0.1, 0.1)$. Define the likelihood $\mathcal{L}(\beta, \theta, X_1, \dots, X_n) = \mathcal{L}(\beta, \theta, \mathcal{X}) = \prod_{i=1}^n \prod_{j=1}^{n_i} p_{ij}(t_{ij}|\beta, \theta, \mathcal{X}_i)^{\delta_{ij}} S_{ij}(t_{ij}|\beta, \theta, \mathcal{X}_i)^{1-\delta_{ij}}$, where $\delta_{ij} = 0$ if observation ij is censored, and 1 if not; t_{ij} is the observed event time. Let the ij^{th} data point be $\mathcal{D}_{ij} = (t_{ij}, \delta_{ij})$ and $\mathcal{D} = \{\mathcal{D}_{ij} : i = 1, \dots, n, j = 1, \dots, n_i\}$. The posterior density

is $p(\beta, \theta, \{\mathcal{X}_i\}, \{\lambda_{l,k}\}|\mathcal{D}) \propto \mathcal{L}(\beta, \theta, \mathcal{X}_1, \dots, \mathcal{X}_n) \prod_{l=2}^M \prod_{k=1}^{2^{l-1}} p(\mathbf{X}_{l,2k-1}|\lambda_{l,2k-1})p(\lambda_{l,2k-1})$. Here, $p(\mathbf{X}_{k,2k-1}|\lambda_{l,2k-1})$ is (6) with $\lambda = \lambda_{2k-1}$ and transformed via (4). The MCMC algorithm for sampling model parameters is outlined in Appendix A.

In data analyses the number of MPT levels was capped at $M = 4$, achieving good MCMC mixing with acceptable computing time. Taking 40000 iterations after a burn-in of 20000 took about 6 hours in compiled FORTRAN 90 on a 1.5 GHz Windows-based PC with 1 GB RAM. We also tried $M = 5$ and found slight deterioration in model performance from the predictive model selection criteria described in Section 3.1.

3. Analysis of the Iowa SEER data

The Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute provides county-level cancer data on an annual basis for public use for particular states. The SEER database provides information on more than 3 million in situ and invasive cancer cases, and the registries routinely collect data on patient demographics, primary tumor site, morphology, stage at diagnosis, first course of treatment, and follow-up for vital status. We apply our proposed model to 1989–1991 SEER breast cancer survival data from the State of Iowa. These data include information on a cohort of 5786 women who were diagnosed with malignant breast between 1989 and 1991, with follow-up continued through the end of 2003. Only deaths due to metastasis of cancerous nodes in the breast were considered to be events, while the rest (including death from metastasis of other types of cancer, or from other causes) were considered to be censored observations. By the end of 2003, 3156 of the patients had died of breast cancer, while the remaining 2630 women were censored because they survived until the end of the study period, died of other causes, or were lost to follow-up. Since the interest of this paper is to develop new methodology to capture the spatial pattern, the temporal effect is not investigated here; i.e. the time-window is chosen to be short enough that the time effect is minimized, and before documented decreasing temporal trend in mortality in Iowa took effect (e.g. Banerjee and Carlin, 2003; Zhao and Hanson, 2008).

For each individual, the dataset records the event time in months, $t_{ij} \in [1, 179]$, and her county of residence at diagnosis, $i = 1, \dots, 99$. The dataset also records age at diagnosis and the stage of the disease. There are three levels of disease stage: local (confined to the breast), regional (spread beyond the breast tissue), or distant (metastasis). We treat “local” as the baseline, and create two dummy variables for “regional” and “distant,” respectively. Thus $\mathbf{z}_{ij}$ is 3-dimensional. Table 2 shows several summary statistics for the data.

3.1 Model comparison

The primary statistic used here is the log pseudo marginal likelihood (LPML), originally suggested by Geisser and Eddy (1979). LPML is defined based on the conditional predictive ordinate (CPO) statistic for the ij^{th} observation. CPO_{ij} is given by $p_{ij}(t_{ij}|\mathcal{D}_{(-ij)})^{\delta_{ij}} S_{ij}(t_{ij}|\mathcal{D}_{(-ij)})^{1-\delta_{ij}}$, the predictive density (for events) or survival function (when censored) of the observed t_{ij} given the remaining data $\mathcal{D}_{(-ij)}$. The LPML is defined as

$\text{LPML} = \log \left(\prod_{i=1}^n \prod_{j=1}^{n_i} \text{CPO}_{ij} \right) = \sum_{i=1}^n \sum_{j=1}^{n_i} \log(\text{CPO}_{ij})$, the log of the product of these conditional density values. Larger LPML indicate “better support” or “greater predictive ability.” Ibrahim, Chen, and Sinha (2001) show that CPO_{ij} is easily estimated by

$\text{CPO}_{ij} = E_{\Omega} \left\{ \frac{1}{p_{ij}(t_{ij}|\Omega)} \right\}$ when $\delta_{ij} = 1$, and $\text{CPO}_{ij} = E_{\Omega} \left\{ \frac{1}{S_{ij}(t_{ij}|\Omega)} \right\}$ otherwise. Here, $\Omega = (\beta, \theta, \mathcal{X}_1, \dots, \mathcal{X}_n)$. The LPML statistic has increased in popularity among Bayesians due to the relative ease of which LPML is stably estimated from MCMC output.

Spiegelhalter et al. (2002) recommend the use of $\bar{D} = -2E_{\Omega} \mathcal{D} \{ \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n) \}$ as a measure of Bayesian model adequacy. The Deviance Information Criterion (DIC; Spiegelhalter et al., 2002), the sum of $\bar{D}$ and a deviance-based model complexity measure, p_D , is commonly used in Bayesian model comparison. Draper and Krnjajic (2007, Sec. 4.1) have shown that DIC approximates the LPML in some situations. We also consider the full Sample Score (FSS), which is the log of the predictive density based on the entire dataset (Laud and Ibrahim 1995; Draper and Krnjajic, 2007). The FSS statistic is computed via

$$\text{FSS} = \sum_{i=1}^n \sum_{j=1}^{n_i} \log E_{\Omega} \{ p_{ij}(t_{ij} | \boldsymbol{\Omega})^{\delta_{ij}} S_{ij}(t_{ij} | \boldsymbol{\Omega})^{1-\delta_{ij}} \}. \text{ Draper and Krnjajic (2007) show that FSS can better estimate expected utility than LPML in certain models.}$$

3.2 Models fit and population effects

We fit the proposed PH model with MPT baselines defined through CAR priors, including models with independent precision parameters $\lambda_{l,k}$ (Model 1) and with common precision parameter λ (Model 1A). We also fit the traditional PH intrinsic CAR (ICAR) frailty model with a common MPT survival baseline (Model 2), proposed by Zhao, Hanson, and Carlin

(2009): $T_{ij} \sim S_{ij}(t | \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \gamma_i) = S_0(t)^{\exp(\mathbf{z}_{ij}'\boldsymbol{\beta} + \gamma_i)}$, $S_0 | \boldsymbol{\theta} \sim PT(c, \rho, G_{\boldsymbol{\theta}})$, $\gamma_i | \lambda \sim ICAR(\lambda)$, where $p(\boldsymbol{\beta}, \boldsymbol{\theta}) \propto 1$ independent of $\lambda \sim \Gamma(0.1, 0.1)$. This model differs from the (non-additive predictor) model in Hennerfeind et al. (2006) only in the prior on $S_0(t)$, which in the latter paper is

given by $S_0(t) = \exp \left[- \int_0^t \exp \left\{ \sum_{j=1}^{K+d-1} b_j \psi_j(s) \right\} ds \right]$, where $\{\psi_j(t)\}_{j=1}^{K+d-1}$ are B-splines of degree d over K knots equispaced over the range of the observed survival times; we used the defaults in BayesX (Brezger, Kneib, and Lang, 2005), a free software package which allows the user to easily implement a variety of spatially structured additive generalized linear mixed models: $d = 3$ and $K = 20$. The vector $\mathbf{b} = (b_1, \dots, b_{K+d-1})'$ has a degenerate prior $p(\mathbf{b} | \tau) \propto \exp(-0.5 \mathbf{b}' \mathbf{P} \mathbf{b})$ where $\mathbf{P}$ is a second order difference penalty matrix (e.g. Kneib, 2006, page 34), completed by the default hyperprior $\tau \sim \Gamma(0.001, 0.001)$.

For Model 1A, Models 1 and 2, the MPT priors on baseline survival are centered at the log-logistic family. We also fit the parametric log-logistic PH model stratified by county (Model

3), i.e. $T_{ij} \sim S_{ij}(t | \boldsymbol{\beta}, \boldsymbol{\theta}_i) = \{1 + e^{\eta_i} t^{\exp(\alpha_i)}\}^{-\exp(\mathbf{z}_{ij}'\boldsymbol{\beta})}$, $p(\boldsymbol{\beta}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{99}) \propto 1$. We fit Models 1 and Model 1A using the algorithm described implemented in Fortran 90. Despite the high dimension of our models, the MCMC chains mixed reasonably well. For Model 1, we retained 40,000 iterations for posterior estimation following a burn-in of 20,000 iterations. For Model 2 and 3, we retained 100,000 iterations for posterior estimation following a burn-in of 50,000 iterations. We also fit the penalized spline model of Hennerfeind, Brezger, and Fahrmeir (2006) as described in Section 4 below. Tables 3 shows DIC, LPML and FSS scores for Models 1, 2, 3 and the spline model.

We tried 4 different values for the propriety parameter r in Model 1; 0.99 performs best according to $\bar{D}$, DIC, LPML and FSS. The propriety parameter r can be viewed as the

expected proportional “reaction” of $Y_{l,k}(i)$ to $\sum_{s \neq i} \frac{\omega_{is}}{\omega_{i+}} Y_{l,k}(s)$ (Banerjee, Carlin, and Gelfand, 2004, p. 81). At $r = 0$, the elements of $\mathbf{Y}_{l,k}$ become independent, and the PH model with spatially independent MPT survival baselines is obtained: a Bayesian analogue to the model fit in SAS PROC PHREG with the STRATA subcommand. Note that Model 3 is another type of stratified model, where each county has a unique independent baseline survival curve. As r tends to 1, the proper CAR prior tends to an intrinsically autoregressive prior. Comparing the models in Model 1 with different r values, DIC, LPML and FSS all show an increase of the goodness of fit as r increases. Comparing across Models 1, 1A, 2 and 3, DIC, LPML and FSS all indicate the same trend for goodness of fit, with the spatially dependent MPT model outperforming the traditional MPT CAR frailty model (the pseudo Bayes

factors are about 90 or 45 for common or independent precision parameters respectively), which outperforms the i.i.d. strata parametric model. Model 2 performs better than Model 1 with $r = 0.9$, while not as good as Model 1 with $r = 0.99$. Model 1 with $r = 0$ (independent) and $r = 0.5$ have similar but the poorest model fit among Model 1 models, and also perform worse than Model 2. Therefore, including the spatial pattern seems important to model prediction and fit. Meanwhile Model 1 with $r = 0$ has better goodness of fit than Model 3, which indicates that MPT part also offers a contribution to improve model fit over a fully parametric alternative. For Model 1, the 95% credible intervals for the model with independent $\lambda_{l,k}$ and $r = 0.99$ are very close. Thus allowing different $\lambda_{l,k}$ for these data is not necessary: Model 1A with $r = 0.99$ and $\lambda_{l,k} = \lambda$ has the best fit among all models according to DIC, LPML and FSS. Comparing the penalized spline model from BayesX, Models 1A, 1 with $r = 0.99$, and 2 have significantly better goodness of fit than the spline model according to DIC. Note that Model 2 is equivalent to the spline model except for the prior on $S_0(t)$, thus indicating that the MPT model better predicts the data within the context of an intrinsic CAR spatial frailty model.

In terms of the induced precisions $c_i = 2\lambda/\sigma_{ii}$ under the Iowa data, $r = 0$ gives $0.14 < \sigma_{ii} < 0.5$; $r = 0.5$ gives $0.15 < \sigma_{ii} < 0.55$; $r = 0.9$ gives $0.21 < \sigma_{ii} < 0.83$. For all of these the smallest value occurs for Webster County with $\omega_{94+} = 7$ neighbors and the largest occurs at a Lyon County, in the northwest corner of Iowa with $\omega_{60+} = 2$ neighbors. When $r = 0.99$ we have $0.43 < \sigma_{ii} < 1.39$. The largest value again occurs for Lyon County, but the smallest shifts to Marion County with $\omega_{63+} = 6$ neighbors. As r tends to unity, the prior mass on c_i builds up near zero, making posterior inference more difficult to obtain (Hanson, 2006).

Table 4 compares posterior medians and equal-tailed 95% credible intervals for main effects (components of β) under the Bayesian models with those obtained under standard semiparametric partial likelihood-based PH. The smoothing spline results are obtained using BayesX and the standard results are obtained using PHREG procedure in SAS 8.2. As is often the case with main effects, point estimates change little across models, while the 95% CI's of the best fitting spatially smoothed model 1A are narrower than those obtained from the stratified model fit in PROC PHREG by 10.7% (distant), 22.9% (regional) and 59.0% (age). All models indicate that the predictors are significant at the 0.05 level. Higher age at diagnosis increases the hazard; e.g., a twenty-year increase in age is associated with a $e^{0.049 \times 20} \approx 2.66$ -fold increase in hazard rate. After adjusting for the age at diagnosis and the county of residence, the hazard rate of dying from breast cancer before any time t is $e^{0.64} \approx 1.90$ times greater for regional stage versus local stage, and is $e^{2.05} \approx 7.77$ times greater for distant stage versus local stage.

3.3 Spatial inferences

Figure 1 offers a geographic summary of the spatial pattern of posterior mean 10th percentiles of survival for women with mean age at diagnosis local stage of disease. These maps show the predicted time in months past which 90% will survive for each county under Models 1, 1A, 2 and 3. We see clusters of poorer survival in the southeast and better survival in north-central to northeast parts of the state. The first three maps are spatially smoothed, and clearly show patterns of breast cancer mortality across the state. Comparing these three maps, the map from Model 1 is more globally smoothed, with best survival rate in north-central and northeast parts, moderate survival in the south-central and southwest parts, and poorest survival in the southeast part of the state. Model 1A and Model 2 produce more "clusters" of survival than Model 1; the spatial smoothing is more localized.

The posterior mean and standard deviation of the 10th and 40th percentiles from Model 1 are plotted in Figure 2; both sets of quantiles are estimated reasonably well relative to their standard deviations. For the 10th percentile, the southeastern part of Iowa fares a bit worse

than the southwestern portion; however this trend is reversed for the 40th percentile. This indicates a gross geographic “shift” in increased survival in terms of percentiles. Taking this further, for the 40th percentile we see a vertical stripe of elevated survival through the middle of the state, perfectly corresponding to Interstate 35, except at Polk county which contains Des Moines. Additionally, elevated survival in the Northwestern part of the state corresponds to Interstate 90 running through the very southern part of Minnesota. Due to the centralization of cancer treatment and expertise, travel time is increasingly a factor in access and utilization of health care services (Payne et al., 2000), affecting mammogram rates (Haynes et al., 2006; Zenk, Tarlov, and Sun, 2006) which can further affect the severity and stage at first diagnosis, breast cancer treatment (Celaya et al., 2006; Jones et al., 2008), and stage at breast cancer diagnosis (Marchick and Henson, 2005; Wang et al., 2008). Proximity to a major highway typically lessens travel time; population density is also correlated to highway access and thus so are clinics (Marchick and Henson, 2005). However, travel time may be overshadowed by other barriers to care in urban settings.

Counties with the worst survival prospects house the cities of Davenport, Waterloo, and Council Bluffs. The part of the Interstate 35 corridor where survival drops roughly 5 months occurs at the Des Moines metropolitan area. Differences in urban versus rural areas in mammography rates were noted by Jackson et al. (2009), adjusting for other variables such as household income and number of nearby mammography facilities. In fact, socioeconomic and cultural barriers can outweigh distance to nearby facilities in preventing mammogram screening in urban environments (Zenk et al., 2006).

The hazard for women with a mean diagnosis age of 65.3 years and “local” stage of breast cancer across Models 1, 1A and 2 is compared for two specific cases. Johnson and Linn are adjacent counties; Johnson and Polk are separated by three counties. Johnson county had 68 events out of 131 diagnoses, Linn county had 160 events out of 312 diagnoses, and Polk county had 315 events out of 638 diagnoses. The upper three plots in Figure 3 provide the predictive hazard curves for the two adjacent counties Johnson and Linn, and the lower three plots are for the two more separated counties, Johnson and Polk. The left two plots show the hazard curves from Model 1, the PH model with spatially dependent MPT baselines and independent precision parameters $\lambda_{l,k}$; the middle two plots are from Model 1A, the PH model with spatially dependent MPT baselines and common precision parameter λ ; and the right two plots report results for Model 2, the traditional CAR frailty PH model. Model 2 plots show similar shapes stretched along the y-axis, reflecting proportional hazards across counties; the hazard ratio between Johnson and Linn is 0.98, between Johnson and Polk it is 0.99. Model 1 and 1A plots show somewhat different shapes; the hazard curves for two counties cross through the follow up time. The plots from Model 1 are essentially smoothed versions of Model 1A from allowing flexible precisions $\lambda_{l,k}$. Furthermore under Model 1, the hazard curves of two adjacent counties are very close, while the curves from the two well-separated counties have quite different shapes. Under Model 2, the hazard ratio for the two neighbor counties and for the well-separated counties are very close: 0.98 and 0.99 respectively. Under Model 1 and 1A, the hazard ratio ranges of the two adjacent counties are (0.97, 1.07) and (0.98, 1.06) respectively, while the the two separated counties are about doubled, (0.87, 1.11) and (0.92, 1.08), respectively.

4. Simulated data

Three illustrative sets of survival times ranging over a simple map were generated and various baseline models compared when the truth is known. Seven counties follow one another in a line yielding the adjacency matrix $\mathbf{W}$ with entries $\omega_{i,i-1} = \omega_{i,i+1} = 1$ for $i = 2, \dots, 6$, $\omega_{1,2} = \omega_{7,6} = 1$, and zero elsewhere. Each county has baseline survival $S_{0i}(t) = 1 - p_i \Phi(t|\mu_{i1}, \sigma_{i1}) - (1 - p_i) \Phi(t|\mu_{i2}, \sigma_{i2})$, a mixture of two normal distributions for the first two

simulations; for the third simulation the baselines follow a conditional proportional hazard model $S_{0i}(t) = S_0(t)^{\gamma_i}$ where $S_0(t)$ is a mixture of two normals. Since the models considered only differ in baseline assumptions, we consider the no-covariates case without any loss of generalization. This has the added benefit of comparing the approaches in terms of density estimation.

CAR MPT models with $M = 4$ and $r = 0.99$ (correlated baselines) and $r = 0$ (purely stratified) are compared to the geoadditive CAR frailty model of Hennerfeind, Brezger, and Fahrmeir (2006). In this latter model the baseline log-hazard function is modeled as a penalized B-spline:

$$\log\{h_0(t)\} = \sum_{j=1}^{K+1} b_j \psi_j(t), \quad (11)$$

where $\{\psi_j(t)\}_{j=1}^{K+1}$ are quadratic B-splines ranging over equispaced knot locations $s_1, \dots, s_K$ spanning the survival times $[t_{\min}, t_{\max}]$; here we take $t_{\min} = 0$ and $t_{\max} = 15$ (see Figure 4). Specifically, $q(u) = 0.5(u^2)I_{[0,1]}(u) + (-3 + 6u - 2u^2)I_{[1,2]}(u) + (3 - u^2)I_{[2,3]}(u)$ is a B-spline over $[0, 3]$ and $\psi_j(t) = q((t - t_{\min})/\Delta + 3 - j)$ where $\Delta = (t_{\max} - t_{\min})/(K - 1)$. The basis coefficients are assigned a first-order random walk prior, $p(\mathbf{b}|\lambda) \propto \lambda^{(K+1)/2} \exp(-0.5\lambda \mathbf{b}'\mathbf{P}\mathbf{b})$ where $\mathbf{P}$ is a $(K + 1) \times (K + 1)$ penalty matrix with entries $p_{11} = p_{K+1,K+1} = 1$, $p_{jj} = 2$ for $j = 2, \dots, K$, and $p_{j,j+1} = p_{j+1,j} = -1$ for $j = 1, \dots, K$. The prior on $\mathbf{b}$ is degenerate normal that penalizes the difference of adjacent coefficients but places an improper flat prior on their overall location. The model is completed through $n = 7$ spatial log-frailties assigned an intrinsic CAR prior, i.e. $\gamma \sim \text{Proper CAR}(\lambda, r)$. The frailty distribution location is confounded with the location of $h_0(t)$, and so a sum-to-zero constraint is enforced for identifiability:

$$\sum_{i=1}^7 \gamma_i = 0. \text{ Conditional proportional hazards yields hazard } h_i(t) = e^{\gamma_i} h_0(t) \text{ for county } i.$$

A member of the editorial board suggested augmenting the model of Hennerfeind et al. (2006) in a manner similar to what we have done with Polya trees. In this spirit, we consider a third model ("CAR B-spline") where log-hazards are correlated over an areal grid through their B-spline coefficients. Specifically, $\mathbf{b}_j = (b_{1j}, b_{2j}, \dots, b_{7j})'$ are coefficients for the j th basis function $\psi_j(t)$ in the 7 counties. We assume $K + 1$ independent proper CAR priors with common precision λ and $r = 0.99$ for each $\mathbf{b}_j$. All precision parameters (λ and τ) are assumed to follow a $\Gamma(0.1, 0.1)$ hyperprior. In this model the i th hazard is an exponentiated B-spline as in (11), but not penalized within each county; rather the coefficients are smoothed or penalized in space, similar to the CAR MPT model. For the Hennerfeind et al. (2006) model and the CAR B-spline model $K = 30$ equispaced knots were chosen over $[0, 15]$.

The first two simulated data sets do not have conditionally proportional hazard functions. The first fixes both normal component means and variances, and changes weights for counties 1, 2, 3, 5, 6, 7; for county 4 the components are closer together with equal weights. The second simply moves equally-weighted components further apart with each adjacent county (see Figure 4). The third simulated data set does exactly follow the frailty model and so should favor the model of Hennerfeind et al. (2006). The data are generated from their distributions in the form of quantiles of equally-spaced probability; each county contributes $n_i = 100$ observations for 700 total. The data sets thus approximate their expected order statistics and are a convenient "perfectly representative" sample. Due to the amount of time required to fit each model, an exhaustive Monte Carlo study with randomly generated data is not feasible with large sample sizes; however, this study gives an idea of how models perform under ideal circumstances.

Each model was fit using standard MCMC updating schemes in compiled FORTRAN 90. A burnin of 2000 was followed by a run of 50,000 iterations, on which posterior inference was based. All models mixed reasonably well. The models are compared for predictive ability using LPML and model fit using the L_1 distance between the true densities $f_i(t)$ and posterior means $\hat{f}_i(t)$ under the models. Tables 5 and 6 show these statistics for the four models over the seven counties, and also the sum over the seven counties as an aggregate measure. For the first two simulations, the CAR MPT model with $r = 0.99$ outpredicts the remaining models, including the CAR MPT model with $r = 0$, a purely stratified model that does not borrow strength across the counties. Allowing correlation across counties significantly improves both prediction and model fit as seen in Tables 5 and 6. However, the CAR B-spline model provides the best fit in terms of L_1 in the first two simulations, as well as providing smoother density estimates. Unfortunately, the CAR B-spline model takes over 5 times as long to obtain the same number of MCMC iterates, due to the necessity of numerical integration to evaluate $S_{0i}(t)$ over a grid of t values for $i = 1, \dots, 7$. Roughly, the CAR MPT model produced 200 iterates per minute, the Hennerfeind et al. (2006) model 150 iterates/minute, and the CAR B-spline model 30 iterates/minute.

The third simulation does follow conditionally proportional hazards. Here the Hennerfeind et al. (2006) model grossly outperforms the other two approaches. Other PH frailty models with flexible baselines (e.g. Zhao et al., 2009) should similarly fare well here. This last simulation shows that if the frailty model holds, a model that structurally imposes this assumption can do much better in terms of both prediction and fit. Otherwise, if district baselines are spatially varying in a manner not accommodated by frailties, the frailty approach can fail miserably (as shown in the center column of Figure 4).

5. Discussion and future work

In this paper, we have developed new methodology to capture the spatial pattern for spatially oriented time-to-event data. The proposed PH model assumes a spatially dependent MPT prior based on logistic-transformed CAR priors on conditional probabilities. Specifically, the logit transformed MPT conditional probabilities follow a proper CAR prior. We compare this spatially dependent MPT approach to the traditional spatial frailty approach in simulations, and illustrate the usefulness of the proposed model with an analysis of 1989–1991 Iowa breast cancer data from SEER program.

As shown in Section 3.2, the proposed method can provide geographic maps of any survival statistic of interest, concisely illustrating how county level survival changes over space. This “big picture” spatial survival pattern, adjusted for covariates, can generate hypotheses concerning omitted geographical risk factors of cancer survival that merit further investigation. The simulations (Section 4) show rather drastic improvement of predictive ability and fit by the proposed CAR MPT model when the structure imposed by the frailty model is violated; in such cases the frailty model can provide markedly misleading results. Of additional interest is the spatially smoothed B-spline model, which holds promise and would be a worthy subject of future exploration. This latter model improved fit as measured by L_1 but reduced predictive ability according to LPML relative to the CAR MPT model.

In CAR MPT model, a dependent MPT prior for each county is assumed, hence $n = 99$ MPT priors across Iowa for the SEER data. We use $M = 4$ as the finite Polya tree level, and so $2^4 - 2 = 14$ conditional probabilities π_i in each MPT prior. For these 14 MPT probabilities, each one corresponds to a proper CAR model with an independent precision parameter. In total, we introduce $(99 + 1) \times 14 = 1400$ parameters, albeit not necessarily ‘free,’ to capture the spatial pattern. If there are enough subjects in each county, this model may show improved predictive power over the traditional frailty approach. While in the case of very

few subjects in each county, although the model adequacy can be improved, a larger penalty may be added to the new approach as a “trade off”. We applied this approach to a smaller dataset with a sample size of $\sum_{i=1}^{99} n_i = 1000$ across the 99 counties. Although $\bar{D}$ and FSS favor the new approach rather than the traditional (frailty) one, the DIC and LPML support the simpler frailty approach better.

To check the robustness of the prior assumptions, we also employed an improper flat prior for (θ', β') . The results indicate the same goodness of fit trend across models. Furthermore, the survival estimates at both population level and geographic subregion level are very close assuming either the proper or improper prior.

In addition to the proper CAR prior, we also considered intrinsic CAR (ICAR) structure for the logit transformed MPT probabilities, i.e., Model 1 with $r \rightarrow 1$. Although among the models in Model 1, the goodness of fit increases as r increases, the ICAR structure did not show an improvement over the proper CAR alternative; the LPML and DIC values from $r \rightarrow 1$ are a bit worse than $r = 0.99$. This may be because that to remedy the non-identifiability of

the ICAR model, a sum-to-zero constraint is commonly employed, $\sum_{i=1}^{99} Y_{l,k}(i) = 0$, for each of the $2^M - 2$ intrinsic CAR distributions. This precludes, for example, having the same conditional probability “bump” refinement to the underlying log-logistic baseline in each of the 99 counties. In particular, it precludes having the same nonparametric baseline across counties, i.e. the model of Zhao et al. (2009), unless $Y_{l,k} = 0$ for all l, k (yielding the parametric log-logistic model). Figure 3 in fact shows roughly similar shapes from 130 to 200 months, but different shapes from 30 to 130 months across three counties. This artificial constraint may not be appropriate for these data, and so model fit does not improve but rather decreases a bit.

Acknowledgments

The authors thank editor Dr. M. Zucker, the associate editor, and the referee for their helpful comments that led to improvement of the article.

References

- Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B.* 1972; 34:187–200.
- Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* New York: Springer; 2001.
- Banerjee S, Carlin BP. Semiparametric spatio-temporal frailty modeling. *Environmetrics.* 2003; 14:523–535.
- Banerjee, S.; Carlin, BP.; Gelfand, AE. *Hierarchical Modeling and Analysis for Spatial Data.* Boca Raton, FL: Chapman and Hall/CRC Press; 2004.
- Banerjee S, Dey DK. Semi-parametric proportional odds models for spatially correlated survival data. *Lifetime Data Analysis.* 2005; 11:175–191. [PubMed: 15938545]
- Banerjee S, Wall MM, Carlin BP. Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics.* 2003; 4:123–142. [PubMed: 12925334]
- Bastos LS, Gamerman D. Dynamic survival models with spatial frailty. *Lifetime Data Analysis.* 2006; 12:441–460. [PubMed: 17031498]
- Benzi M, Golub GH. Bounds for the entries of matrix functions with applications to preconditioning. *Bit Numerical Mathematics.* 1999; 39:417–438.
- Besag J. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B.* 1974; 36:192–236.
- Besag J, Kooperberg C. On conditional and intrinsic autoregressions. *Biometrika.* 1995; 82:733–746.

- Brezger A, Kneib T, Lang S. BayesX: Analyzing Bayesian structured additive regression models. *Journal of Statistical Software*. 2005; 14:1–22.
- Celaya MO, Rees JR, Gibson JJ, Riddle BL, Greenberg ER. Travel distance and season of diagnosis affect treatment choices for women with early-stage breast cancer in a predominantly rural population (United States). *Cancer Causes Control*. 2006; 17:851–856. [PubMed: 16783613]
- Draper, D.; Krnjajic, M. Technical report. Department of Applied Mathematics and Statistics, University of California; Santa Cruz: 2007. Bayesian model specification.
- Dunson DB, Pillai N, Park JH. Bayesian density regression. *Journal of the Royal Statistical Society, Series B*. 2007; 69:163–183.
- Ferguson TS. Prior distributions on spaces of probability measures. *The Annals of Statistics*. 1974; 2:615–629.
- Geisser S, Eddy WF. A predictive approach to model selection. *Journal of the American Statistical Association*. 1979; 74:153–160.
- Greenland S, Christensen R. Data-augmentation priors for Bayesian and semi-Bayes analysis of conditional-logistic and proportional-hazards regression. *Statistics in Medicine*. 2001; 20:2421–2428. [PubMed: 11512132]
- Griffin JE, Steel MF. Order-based dependent Dirichlet processes. *Journal of the American Statistical Association*. 2006; 101:179–194.
- Hanson T, Johnson WO. Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association*. 2002; 97:1020–1033.
- Hanson T. Inference for mixtures of finite Polya tree models. *Journal of the American Statistical Association*. 2006; 101:1548–1565.
- Hanson T, Bedrick EJ, Johnson WO, Thurmond MC. Modeling fetal survival in dairy cattle. *Statistics in Medicine*. 2003; 22:1725–1739. [PubMed: 12720307]
- Haynes R, Jones AP, Sauerzapf V, Zhao H. Validation of travel times to hospital estimated by GIS. *International Journal of Health Geographics*. 2006; 5:1–8. [PubMed: 16390549]
- Henderson R, Shimakura S, Gorst D. Modeling spatial variation in leukemia survival data. *Journal of the American Statistical Association*. 2002; 97:965–972.
- Hennerfeind A, Brezger A, Fahrmeir L. Geoadditive survival models. *Journal of the American Statistical Association*. 2006; 101:1065–1075.
- Ibrahim, JG.; Chen, M-H.; Sinha, D. *Bayesian Survival Analysis*. New York: Springer; 2001.
- Jackson MC, Davis WW, Waldron W, McNeel TS, Pfeiffer R, Breen N. Impact of geography on mammography use in California. *Cancer Causes Control*. 2009; 20:1339–1353. [PubMed: 19449107]
- Jones AP, Haynes R, Sauerzapf V, Crawford SM, Zhao H, Forman D. Travel time to hospital and treatment for breast, colon, rectum, lung, ovary and prostate cancer. *European Journal of Cancer*. 2008; 44:992–999. [PubMed: 18375117]
- Kneib, T. PhD Thesis. Munich University; 2006. Mixed Model Based Inference in Structured Additive Regression.
- Komárek A, Lesaffre E, Legrand C. Baseline and treatment effect heterogeneity for survival times between centers using a random effects accelerated failure time model with flexible error distribution. *Statistics in Medicine*. 2007; 26:5457–5472. [PubMed: 17910009]
- Kottas A, Duan JA, Gelfand AE. Modeling disease incidence data with spatial and spatio-temporal Dirichlet process mixtures. *Biometrical Journal*. 2008; 50:29–42. [PubMed: 17926327]
- Laud P, Ibrahim J. Predictive model selection. *Journal of the Royal Statistical Society, Series B*. 1995; 57:247–262.
- Lavine M. Some aspects of Polya tree distributions for statistical modeling. *Annals of Statistics*. 1992; 20:1222–1235.
- Lavine M. More aspects of Polya tree distributions for statistical modeling. *Annals of Statistics*. 1994; 22:1161–1176.
- Li Y, Ryan L. Modeling spatial survival data using semiparametric frailty models. *Biometrics*. 2004; 58:287–297. [PubMed: 12071401]

- Marchick J, Henson DE. Correlations between access to mammography and breast cancer stage at diagnosis. *Cancer*. 2005; 103:1571–1580. [PubMed: 15772962]
- Payne S, Jarrett N, Jeffs D. The impact of travel on cancer patients' experiences of treatment: A literature review. *European Journal of Cancer Care*. 2000; 9:197–203. [PubMed: 11829366]
- Reich BJ, Fuentes M. A multivariate nonparametric Bayesian spatial framework for hurricane surface wind fields. *The Annals of Applied Statistics*. 2007; 1:249–264.
- Spiegelhalter DJ, Best N, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*. 2002; 64:583–639.
- Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*. 1979; 16:439–454. [PubMed: 510638]
- Walker SG, Mallick BK. Hierarchical generalized linear models and frailty models with Bayesian nonparametric mixing. *Journal of the Royal Statistical Society, Series B*. 1997; 59:845–860.
- Walker SG, Mallick BK. Semiparametric accelerated life time model. *Biometrics*. 1999; 55:477–483. [PubMed: 11318203]
- Wang F, McLafferty S, Escamilla V, Luo L. Late-stage breast cancer diagnosis and health care access in Illinois. *The Professional Geographer*. 2008; 60:54–69. [PubMed: 18458760]
- Zeng D, Lin DY, Yin G. Maximum likelihood estimation for the proportional odds model with random effects. *Journal of the American Statistical Association*. 2005; 100:470–483.
- Zenk SN, Tarlov E, Sun J. Spatial equity in facilities providing low- and no-fee screening mammography in Chicago neighborhoods. *Journal of Urban Health*. 2006; 83:195–210. [PubMed: 16736369]
- Zhao L, Hanson T, Carlin BP. Mixtures of Polya trees for flexible spatial frailty survival modeling. *Biometrika*. 2009; 96:263–276. [PubMed: 19779579]
- Zhao, L.; Hanson, T. Technical Report. University of Minnesota Division of Biostatistics; 2008. Temporally dependent baseline survival in the proportional hazards model.

Figure 1. Estimated 10th percentiles of survival (months) for women with mean diagnosis age and local stage of disease, Models 1, 1A, 2 and 3.

Figure 2. Posterior mean and standard deviation of 10th and 40th percentiles of survival (months) for women with mean diagnosis age and local stage of disease, Model 1.

Figure 3. Fitted predictive hazard of women with mean diagnosis age and local stage of breast cancer from Models 1, 1A, and 2.

Figure 4. True and estimated densities. Columns 1, 2, 3 are the simulations and rows 1 – 7 are the counties. Thick solid is the truth, thin solid CAR MPT, dashed HBF-frailty, and dots are CAR B-spline.

Table 1

Schematic to define G_{θ} for $l = 1, 2, 3$.

$\mathbf{R}^+$			
$B_{\theta}(1, 1)$		$B_{\theta}(1, 2)$	
$(X_{1,1}, X_{1,2}) = (0.5, 0.5)$			
$B_{\theta}(2, 1)$	$B_{\theta}(2, 2)$	$B_{\theta}(2, 3)$	$B_{\theta}(2, 4)$
$(X_{2,1}, X_{2,2}) \sim \text{Dir}(cp(2), cp(2))$		$(X_{2,3}, X_{2,4}) \sim \text{Dir}(cp(2), cp(2))$	
$B_{\theta}(3, 1)$	$B_{\theta}(3, 2)$	$B_{\theta}(3, 3)$	$B_{\theta}(3, 4)$
$(X_{3,1}, X_{3,2}) \sim \text{Dir}(cp(3), cp(3))$	$(X_{3,3}, X_{3,4}) \sim \text{Dir}(cp(3), cp(3))$	$B_{\theta}(3, 5)$	$B_{\theta}(3, 6)$
$(X_{3,5}, X_{3,6}) \sim \text{Dir}(cp(3), cp(3))$		$B_{\theta}(3, 7)$	$B_{\theta}(3, 8)$
$(X_{3,7}, X_{3,8}) \sim \text{Dir}(cp(3), cp(3))$			

Table 2

Summary statistics for follow-up time and covariates, 1989–1991 Iowa SEER breast cancer data.

continuous variables		mean	std deviation
follow-up time (months)		110.3	57.2
age (years)		65.3	14.7
categorical variables		count	proportion (%)
status	event	3156	54.5
	censored	2630	45.5
stage	local	3618	62.5
	regional	1813	31.3
	distant	355	6.1

Table 3

Expected deviance ($\bar{D}$), DIC, LPML and FSS for the three models and smoothing spline model.

Model	λ	r	$\bar{D}$	DIC	LPML	FSS
1A	common	$r = 0.99$	37823.1	37865.3	-18942.3	-18881.5
1	i.i.d.	$r = 0.99$	37844.5	37876.4	-18943.0	-18902.1
		$r = 0.9$	37866.3	37894.6	-18950.0	-18916.7
	i.i.d.	$r = 0.5$	37885.3	37907.0	-18955.8	-18929.8
		$r = 0$	37885.3	37907.8	-18956.1	-18929.5
2	-	-	37854.2	37880.9	-18946.8	-18908.4
3	-	-	37866.2	38076.8	-19049.1	-18933.6
HBF-frailty	-	-	37855.7	37893.3	-	-

Table 4

Point estimates and 95% equal-tail credible intervals, standard and Bayesian model fixed effects.

Model	β_1 (centered age)	β_2 (regional stage)	β_3 (distant stage)
1A	0.049 (0.048, 0.051)	0.64 (0.59, 0.71)	2.05 (1.92, 2.16)
1 ($r = 0.99$)	0.050 (0.047, 0.053)	0.64 (0.57, 0.70)	2.03 (1.92, 2.16)
2	0.050 (0.047, 0.053)	0.64 (0.57, 0.71)	2.05 (1.92, 2.16)
3	0.049 (0.046, 0.052)	0.63 (0.55, 0.70)	2.09 (1.96, 2.21)
HBf-frailty by BayesX	0.049 (0.046, 0.053)	0.62 (0.55, 0.70)	2.02 (1.90, 2.13)
Standard PH model by SAS	0.049 (0.046, 0.052)	0.64 (0.56, 0.72)	2.06 (1.94, 2.19)

Table 5

Simulated data: negative LPML (smaller is better).

Sim	Model	County						
		1	2	3	4	5	6	7
1	Model 1A $r = 0.99$	178.8	206.8	215.2	209.9	215.5	207.3	179.8
	Model 1A $r = 0.0$	182.8	213.1	222.0	204.4	222.1	214.5	185.2
	HBf-frailty	183.8	209.5	212.1	218.2	211.9	203.8	180.8
	CAR B-spline	181.8	208.8	217.9	206.9	218.3	208.9	182.9
2	Model 1A $r = 0.99$	147.6	157.4	180.1	200.4	211.9	216.9	222.4
	Model 1A $r = 0.0$	151.4	162.6	185.3	205.1	217.0	223.1	226.9
	HBf-frailty	177.0	179.2	186.4	198.6	216.1	239.7	270.7
	CAR B-spline	148.1	158.1	181.8	201.9	213.1	217.9	223.6
3	Model 1A $r = 0.99$	219.0	210.6	205.2	199.4	191.9	181.9	170.4
	Model 1A $r = 0.0$	221.6	215.6	209.9	203.3	195.0	184.3	170.5
	HBf-frailty	211.6	205.2	200.0	194.0	186.2	176.0	162.9
	CAR B-spline	221.6	212.9	207.2	200.8	192.5	181.9	170.0

Table 6

Simulated data: $L_{1i} = \int_0^\infty |\widehat{f}_i(t) - f_i(t)| dt$ (smaller is better), $i = 1, \dots, 7$.

Sim	Model	County						
		1	2	3	4	5	6	7
1	Model 1A $r = 0.99$	0.15	0.14	0.18	0.33	0.19	0.12	0.10
	Model 1A $r = 0.0$	0.18	0.20	0.18	0.12	0.18	0.17	0.12
	HBFB-frailty	0.28	0.32	0.19	0.55	0.18	0.21	0.51
	CAR B-spline	0.07	0.05	0.13	0.22	0.11	0.05	0.16
2	Model 1A $r = 0.99$	0.07	0.06	0.04	0.07	0.09	0.12	0.22
	Model 1A $r = 0.0$	0.07	0.10	0.09	0.13	0.17	0.21	0.25
	HBFB-frailty	0.68	0.58	0.34	0.22	0.38	0.61	0.80
	CAR B-spline	0.06	0.05	0.04	0.04	0.04	0.04	0.11
3	Model 1A $r = 0.99$	0.12	0.08	0.08	0.09	0.10	0.12	0.18
	Model 1A $r = 0.0$	0.22	0.25	0.22	0.21	0.20	0.19	0.16
	HBFB-frailty	0.02	0.02	0.02	0.01	0.02	0.02	0.03
	CAR B-spline	0.07	0.03	0.03	0.03	0.02	0.03	0.04