Published in final edited form as:

Environmetrics. 2018 ; 29(5-6): . doi:10.1002/env.2478.

Bayesian inference in time-varying additive hazards models with applications to disease mapping

A. Chernoukhov^{a,*}, A. Hussein^{b,*}, S. Nkurunziza^{b,*}, and D. Bandyopadhyay^c

^aSenior Risk Analyst, Royal Bank of Canada, Canada

^bDepartment of Mathematics and Statistics, University of Windsor, Canada

^cDepartment of Biostatistics, Virginia Commonwealth University, Richmond, VA, USA

Summary

Environmental health and disease mapping studies are often concerned with the evaluation of the combined effect of various socio-demographic and behavioral factors, and environmental exposures on time-to-events of interest, such as death of individuals, organisms or plants. In such studies, estimation of the hazard function is often of interest. In addition to known explanatory variables, the hazard function may be subject to spatial/geographical variations, such that proximally located regions may experience similar hazards than regions that are distantly located. A popular approach for handling this type of spatially-correlated time-to-event data is the Cox's Proportional Hazards (PH) regression model with spatial frailties. However, the PH assumption poses a major practical challenge, as it entails that the effects of the various explanatory variables remain constant over time. This assumption is often unrealistic, for instance, in studies with long follow-ups where the effects of some exposures on the hazard may vary drastically over time. Our goal in this paper is to offer a flexible semiparametric additive hazards model (AH) with spatial frailties. Our proposed model allows both the frailties as well as the regression coefficients to be time-varying, thus relaxing the proportionality assumption. Our estimation framework is Bayesian, powered by carefully tailored posterior sampling strategies via Markov chain Monte Carlo techniques. We apply the model to a dataset on prostate cancer survival from the US state of Louisiana to illustrate its advantages.

Keywords

Additive hazards; Bayesian; Conditionally autoregressive prior; Metropolis-Hastings; Proposal density; Spatial; Prostate Cancer

1. INTRODUCTION

Analysis of time-to-event data is quite common in environmental studies, where the effect of various environmental exposures on the survival of individuals, plants, micro-organism etc. is of interest. For example, one may be interested in studying the risk of developing certain

*Correspondence to: A. Hussein, Department of Mathematics and Statistics, University of Windsor, 401 Sunset Avenue, Windsor, Ontario N9B 3P4, Canada; ahussein@uwindsor.ca.

types of cancers for subjects living in the vicinity of a nuclear power plant (Patick *et al.*, 2014), or in assessing environmental factors on the risk of suicide (Knox *et al.*, 2013), or in the survival of plant species (Li *et al.*, 2015). Time-to-event outcomes are often censored, whereby the actual times of events are not observable for some of the subjects in the study due to, for example, loss to follow up. In addition, the outcome of interest may also exhibit geographical variations, which manifest in the form of spatial correlations. Popular approaches to the analysis of this type of data are the Cox's Proportional Hazards (PH) (Cox, 1972), the Accelerated Failure Time (AFT) (Zhang and Lawson, 2011) and the Aalen's Additive Hazards (AH) (Aalen, 1980, 1989) regression models with spatial frailties.

The PH model is, by far, the most popular in this class. It models the hazard function of the event of interest and its regression coefficients are readily interpretable as relative risks. Despite its popularity, the Cox's model may not be practical in many situations due to the restrictive proportionality assumption. In particular, the proportionality assumption entails that the effects of the various regressors (e.g., environmental exposures) on the hazard of developing the event of interest remains constant over time. Such an assumption is not reasonable in studies with long follow-up times where the effect of some exposure variables on the hazard may change substantially over time. For example, by using data on breast cancer, Bellera *et al.* (2010) showed that some prognostic factors can have negative effects on the hazard of developing the event of interest early on during follow-up, while later on, such effects may start to reverse. This perhaps embodies a worst-case scenario in violating the PH assumption, where the opposite effects of the exposure cancels out. If a Cox's PH model is still adopted for these situations, the consequences are quite serious as some important exposure variables may be missed all together. On the other hand, the AFT and the AH models offer attractive alternatives to Cox's PH model when the PH assumption is not tenable. The AFT model however, is a fully parametric model, which relies heavily on the correct specification of a parametric distribution (e.g., Weibull, Exponential, and Gompertz). Furthermore, the AFT model's regression coefficients cannot be easily interpreted in terms of hazard (risk) of occurrence of the events of interest, as is the case with the PH and AH models. This leaves us with the AH model, which is a flexible semi-parametric model with time-varying covariate effects that can be easily interpreted in terms of absolute changes in hazard for unit changes in covariate values. This feature represents also an advantage of the AH models over their multiplicative hazards counterparts (e.g., Cox's PH model) in environmental applications, where decision-making is usually based on the expected number of additional cases due to an exposure above and beyond a population's expected baseline number of deaths in the absence of the exposure. For instance, assuming that an exposure variable is a binary indicator, the cumulative regression coefficient corresponding to that exposure, when everything else is held fixed, can be interpreted as the expected additional cumulative number of deaths due to the exposure. Interested reader will find more than ten reasons why an AH model is preferable in some applications than the popular Cox's PH model (Odd *et al.* (2008); p.155–157).

The statistical methodologies for analysis of spatially referenced survival data has grown rapidly in the past decade. For instance, Banerjee *et al.* (2003) developed a Bayesian version of the Cox's PH model with spatial frailties and applied it to infant mortality data, while Banerjee, Lin and Dey (2005) extended it to a proportional odds model. A Cox's model

where splines were used to model effects of piece-wise continuous covariates (e.g., age) and Markov random fields to model spatial effects was proposed in Kneib (2006). This model, although more general than Banerjee et al. (2003), does not accommodate time-varying spatial effects. In the parametric direction, Zhang and Lawson (2011) considered a Bayesian AFT model and used Gaussian frailties to handle spatial dependencies in a disease mapping framework. In a similar vein, Lin (2012) studied an AFT model under an estimating equation approach. Also, Silva and Amaral-Turkman (2004) and Lin and Zhu (2012) studied AH models with frailties, respectively from frequentist and Bayesian points of view. As far as we are aware of, there is no literature that considered an additive hazards model in which both covariate effects and frailties are modeled as time-varying functions.

In this paper, we develop a very general and flexible Bayesian formulation of the Aalen's AH model to handle spatially-correlated survival data by allowing the baseline hazard function, the regression coefficients, as well as the spatial frailties to be time-varying processes. Assuming a piece-wise constant structure for all time-varying functions and spatial frailties following a conditional autoregressive (CAR) structure (see, for instance, Cressie and Wikle, 2006), the computational framework of our additive hazards CAR (henceforth, AH-CAR) model will be carried out via Markov chain Monte Carlo (MCMC) techniques (Gelman et. al, 2014). An appealing feature of this model is allowing both the spatial and the covariate effects to be time-varying functions, thus capturing any changes of exposure, or spatial, effects that may occur over long periods of follow-up. A price to pay for the additional flexibility is often a higher complexity of the model, which can be alleviated by using a Bayesian approach. This was, in fact, the main reason for choosing a Bayesian approach in the current work. We illustrate our methodology by application to a prostate cancer (PrCA) diagnosis dataset for patients from the US state of Louisiana, available in the Surveillance, Epidemiology, and End Results (SEER) database (Surveillance, Epidemiology, and End Results, SEER).

The remainder of the paper is organized as follows. In Section 2, we set up some notations, formulate the AH model with general spatial effects and provide its likelihood function. In Section 3, we provide the Bayesian treatment, where the spatial frailties are conditionally autoregressive. We also specify priors for the rest of the parameters of the model and sketch a framework for carrying out estimation via MCMC. Section 4 applies the model to the motivating SEER PrCA dataset. Finally, Section 5 concludes.

2. AN ADDITIVE HAZARDS (AH) MODEL WITH SPATIAL EFFECTS

It is often convenient to formulate survival models under the general framework of counting processes by defining right-censored survival data collected from N independent subjects as the triplets $\mathbf{D} = \{(N_i(t), Y_i(t), \mathbf{z}_i(t)), i = 1, \dots, N, 0 \leq t \leq \tau\}$, where $\{N_i(t), t \in [0, \tau]\}$ is a process counting the number of occurrences of the event of interest in the time interval $(0, t]$, while $\{Y_i(t), t \in [0, \tau]\}$ is the corresponding at-risk process, which takes a value 1 or 0 according to whether the i^{th} individual is at risk just prior to time t and N is the total number of individuals in the study. The process $\{\mathbf{z}_i(t) = (z_{i1}(t), \dots, z_{ip}(t))^T, \mathbf{z}_i(t) \in \mathbf{\Omega}, i = 1, \dots, N, t \in [0, \tau]\}$ represents a p -dimensional vector of time dependent covariates, where $\mathbf{x}^T$ denotes the transpose of $\mathbf{x}$, and $\mathbf{\Omega} \subset \mathbb{R}^p$ is the set of all admissible covariate vectors (in a sense to be

defined later). In the case of single event per subject (e.g., death), the process $N(t)$ is zero until the event of interest occurs and then it jumps to 1 thereafter while $Y(t)$ serves as censoring indicator. We propose to model the i^{th} individual's hazard of having the event of interest at time t as:

$$h_i(t) = \lambda(t) + \sum_{k=1}^p \alpha_k(t) z_{ik}(t) + \omega_{l_i}(t), \quad 0 \leq t \leq \tau, \quad (1)$$

where $\lambda(t)$ is the baseline hazard, $\alpha_k(t)$, $k = 1, \dots, p$ are time dependent regression coefficients, and $\omega_{l_i}(t)$ is a random and time dependent region-specific frailty term for the region $l_i \in \{1, \dots, n\}$ to which the i^{th} individual belongs. This model is an extension of the usual AH model by including additive, region specific and time-varying effects, $\omega_{l_i}(t)$. In the absence of the spatial effects, this model reduces to the original Aalen's additive hazards model. An issue with additive hazards models is that the estimated hazard function can legitimately take on negative values if the model parameters are allowed to take on all possible real numbers. However, a negative hazard is not meaningful. One approach to circumvent this difficulty is to choose prior distributions such that all parameters in the hazard function are non-negative (Silva and Amaral-Turkman, 2004). This is obviously a restrictive assumption, since, in many situations, the covariates may have structural negative values. Another approach proposed in Cai and Zeng (2011) estimates all the parameters without paying attention to the issue of negativity of the hazard function, but subsequently modifies the estimator of the survival function in such way that it becomes always non-increasing. Cai and Zeng (2011) mentioned that if the non-modified estimator of the survival function is consistent, the modified estimator will also be consistent. While this approach ensures that the survival function estimator is non-increasing function, the actual estimators of the coefficients of the additive hazards are not interpretable when the hazard is negative. In this work we take a flexible approach by constraining the joint distribution of $\lambda(t)$, $\alpha(t)$ and $\omega(t)$ to a region where the cumulative hazard is non-negative for all admissible covariate vectors in Ω . Such constraint can be expressed as

$$\lambda(t) + \sum_{k=1}^p \min \left\{ \alpha_k(t) \inf_{\Omega_k} \alpha_k(t) \sup_{\Omega_k} \alpha_k(t) \right\} + \min_{1 \leq l \leq n} \{ \omega_l(t) \} \geq 0, \quad \forall t \in (0, \tau], \quad (2)$$

where Ω is expressed as a Cartesian product of p sets $\Omega = \Omega_1 \times \dots \times \Omega_p$. This condition will then be incorporated in the joint distribution of the parameters which will be introduced later in our Bayesian framework.

We assume that all time-varying quantities in the above additive hazards are piecewise constant functions over the time partition $(t_0, t_1], (t_1, t_2], \dots, (t_{m-1}, t_m]$, where $t_0, \dots, t_m$ is a finite set of time points such that $0 = t_0 < t_1 < \dots < t_m = \tau$ with $t_j = t_j - t_{j-1}$ being the length of the j th interval. Therefore, in the j th interval, define $\lambda_j \equiv \lambda(t_j)$, $\alpha_{kj} \equiv \alpha_k(t_j)$, $z_{ikj} \equiv z_{ik}(t_j)$, and $\omega_{lj} \equiv \omega_l(t_j)$, to be the values of the piecewise constant functions representing the model

parameters as well as the covariate processes. These pieces can be collected in vectors as λ , α , z , ω , for $i = 1, \dots, N$, $j = 1, \dots, m$, $k = 1, \dots, p$ and $l = 1, \dots, n$. In practice, one can assume equidistant time points with $t_j = t_{j-1} + \Delta t_j$ for all j , and hence, choosing the number of intervals is equivalent to choosing the parameter m , which can be selected by using the various Bayesian model selection criteria available in the literature. The choice of m will be discussed later in Section 4.

Furthermore, the likelihood of the data, given our hazard function formulated above, can be written as follows:

$$L(\mathbf{D}|\lambda, \alpha, \omega) \propto \prod_{j=1}^m \left[\left(\prod_{i \in \mathcal{E}_j} h_i(t_j) \right) \exp \left(- \sum_{i=1}^N R_{ij} h_i(t_j) \Delta t_j \right) \right], \quad (3)$$

where $\mathcal{E}_j = \{i: N_i(t_j) - N_i(t_{j-1}) = 1\}$ denotes the set of all individuals with events in the

interval $(t_{j-1}, t_j]$, and $R_{ij} = \int_{t_{j-1}}^{t_j} Y_i(t) dt / \Delta t_j$ is the proportion of time the i^{th} individual is at risk in the interval $(t_{j-1}, t_j]$.

3. THE BAYESIAN AH MODEL WITH CONDITIONAL AUTOREGRESSIVE FRAILTIES (AH-CAR)

At this point, we consider the AH model defined in the previous section and endow it with random frailties. This is achieved by imposing conditional autoregressive (CAR) priors on the spatial frailties ω_i . Furthermore, in this section we develop priors for all the parameters in the AH model as well as for any hyperparameters. Finally, we provide the form of the joint posterior for all parameters in the model given the data, we then sketch the implementation of a Metropolis-within-Gibbs sampler tailored to estimate the AH-CAR model parameters.

3.1. Choice of the priors and the posterior likelihood

We assume that the prior distributions of the values of the baseline hazard λ_j , $j = 1, \dots, m$ in the partitioning are independent Gamma with shape and scale parameters $a = r_0 c_0 \Delta t_j$ and $b = 1/(c_0 \Delta t_j)$, respectively, with $r_0 > 0$, $c_0 > 0$, and joint distribution

$\pi(\lambda) \propto \prod_{j=1}^m \lambda_j^{r_0 c_0 \Delta t_j - 1} \exp(-\lambda_j c_0 \Delta t_j)$, $\lambda_j > 0$, $\forall j = 1, \dots, m$. The value r_0 represents the ‘best guess’ for the baseline hazard at each interval, and the value c_0 represents the ‘confidence’ in this guess. Such interpretation of r_0 and c_0 follows from the fact that mean of the above Gamma distribution is r_0 and the variance is $r_0/(c_0 \Delta t_j)$, which decreases when c_0 increases. If the baseline hazard is not known, one should set c_0 very small to make the prior vague. As to the regression coefficients, we put flat (improper uniform) priors, so that $\pi(\alpha) \propto 1$.

To account for the spatial association based on the (imposed) adjacency structure of the regions, we consider an (intrinsic) conditional autoregressive (CAR) prior (Banerjee et al., 2003; Zhang and Lawson, 2011), given as:

$$\pi(\omega_j | \theta_j^2) \propto (\theta_j^2)^{-(n-1)/2} \exp \left(- \sum_{\substack{l \sim l' \\ l < l'}} (\omega_{lj} - \omega_{l'j})^2 / (2\theta_j^2) \right) \delta(\omega_{1j}), \text{ where } l \sim l' \text{ denotes the}$$

adjacency relation between regions l and l' with the condition $l < l'$ ensuring that each pair of adjacent regions is included in the summation only once, and θ_j^2 is the variance component, while the indicator function ensures that the first frailty term ω_{1j} is set to zero to safeguard identifiability. The power of θ_j^2 is $(n-1)/2$, since one frailty is fixed and only the remaining $n-1$ frailties are included in the intrinsic CAR specification. This prior was used in Besag et al. (1991) under the name of Gaussian intrinsic autoregression, mainly in applications to image restoration. The latter authors note that such priors are improper since the corresponding variables are of an arbitrary level, and only their differences are taken into account. But this impropriety is removed from the posterior distribution by the presence of informative data.

Finally, we model the hyper-parameters $\theta = (\theta_1, \dots, \theta_m)$ by using independent Inverse-Gamma (conjugate) priors, $\mathcal{IG}(\beta, \gamma)$, with mean $\beta/(\gamma-1)$ and variance $\beta^2/[(\gamma-1)^2(\gamma-2)]$.

Summing up, the joint (constrained) prior of all parameters can be obtained by multiplying all components introduced earlier and an indicator function representing the positivity constraint which was imposed in the previous section:

$$\begin{aligned}
\pi(\lambda, \alpha, \omega, \theta) &\propto \prod_{j=1}^m \underbrace{\left(\lambda_j^{c_0 r_0 \Delta t_j - 1} \exp(-c_0 \lambda_j \Delta t_j) \right)}_{\text{Gamma prior for } \lambda_j} \\
&\times \prod_{j=1}^m \underbrace{\left(\frac{1}{(\theta_j^2)^{(n-1)/2}} \exp \left(-\frac{1}{2\theta_j^2} \sum_{\substack{l \sim l' \\ l < l'}} (\omega_{lj} - \omega_{l'j})^2 \right) \delta(\omega_{1j}) \right)}_{\text{Conditional autoregressive prior for } \omega_j = (\omega_{1j}, \dots, \omega_{nj})^T} \\
&\times \prod_{j=1}^m \underbrace{\left(\mathbb{I} \left\{ \lambda_j + \sum_{k=1}^p \min \left\{ \alpha_{kj} \inf \Omega_k, \alpha_{kj} \sup \Omega_k \right\} + \min_{1 \leq l \leq n} \omega_{lj} \geq 0 \right\} \right)}_{\text{Constraint component}} \times \prod_{j=1}^m \\
&\underbrace{\left((\theta_j^2)^{-\gamma-1} \exp \left(-\frac{\beta}{\theta_j^2} \right) \right)}_{\text{Inverse-gamma prior for } \theta_j^2},
\end{aligned} \tag{4}$$

where $\mathbb{I}\{E\}$ denotes the indicator function of the event E .

The posterior distribution $\pi(\lambda, \alpha, \omega, \theta | \mathbf{D})$ (up to a normalizing constant) is obtained by multiplying (4) and (3), and is given as:

$$\pi(\lambda, \alpha, \omega, \theta | \mathbf{D}) \propto L(\mathbf{D} | \lambda, \alpha, \omega) \pi(\lambda, \alpha, \omega, \theta). \quad (5)$$

3.2. MCMC implementation

In this section, we sketch the computational framework of our AH-CAR model while all technical results are available on the journal's website as supplementary material. Alternatively, one can refer to Chernoukhova (2013) for detailed technical proofs.

Our computations rely on the 'Metropolis-within-Gibbs' MCMC algorithm to sample from the joint posterior distribution given in equation (5). The basic idea of all MCMC algorithms is to construct a Markov chain with the limiting distribution equal to the desired posterior distribution. Metropolis-within-Gibbs, particularly, updates the parameters one-by-one using the full conditional distributions. If the full conditional distributions are not standard, the sampling is performed from the so-called proposal distributions instead of the actual conditionals and the algorithm adjusts for the differences in these distributions by itself in order to obtain the correct limiting distribution.

In our AH-CAR model developed earlier, the parameters of interest for sampling are the baseline hazard λ_j , the regression functions α_{kj} , and the hyper-parameters θ_j that appear in the CAR structure. The frailties ω_{lj} can be considered either as the parameters of interest along with the previous ones or as nuisance parameters. The nuisance parameters are usually integrated out from the joint distribution. However, in our case, such integration is very difficult to carry out, so we sample the frailties along with all other parameters. The full set of parameters to be sampled consists of m parameters for the baseline hazard $\{\lambda_j\}_{j=1}^m$, m parameters for each of p regression functions $\{\{\alpha_{kj}\}_{k=1}^p\}_{j=1}^m$, m parameters for each of the $n-1$ frailty terms $\{\{\omega_{lj}\}_{l=1}^n\}_{j=1}^m$, and m hyper-parameters $\{\theta_j\}_{j=1}^m$ for CAR model, which form $m(1+p+n)$ parameters in total. For MCMC algorithm to work better, on each step, one must ensure that the proposal density is close to the conditional density or at least similar in shape. In order to facilitate a smooth reading of the article, we collected all technical details concerning the derivations of the conditionals of the various parameters as well as derivations of the appropriate proposal densities in the supplementary material document, available at the journal's web site.

The routine MCMC sampling part is implemented in the `C` programming language. The necessary functions are written in `C` and compiled into a shared library. In order to allow analysis of data with very large number of observations (greater than 50,000), all the routine functions use the multiple-precision floating point types from the open source `C` libraries `GMP` and `MPFR`. These libraries allow operations on numbers which are much closer to zero than standard floating-point types of `R` and `C` languages. This is essential for data with large number of observations, since the expressions for the posterior full conditional distributions become extremely small in this case. As a result, the computed expressions become zero due to rounding. The rest of the program is written using the `R` language, which calls the routine MCMC-sampling functions from the compiled `C` shared library.

4. APPLICATION: PRCA SURVIVAL DATA

4.1. Data description and Motivation

In this section, we apply the proposed methodology to the SEER PrCA data mentioned earlier. We re-analyze the dataset from Zhang and Lawson (2011), focussing only on the Louisiana registry (out of 17 SEER registries) having 64 counties, and with a relatively larger proportion of African-American males (29.34%). The prime motivation here is to investigate large geographical variations and racial disparities in the PrCA survival, and to understand the time-varying effects of a number of model covariables. Observations with missing values on the covariates such as Race, Age, County, Stage of PrCA, and Marital status at diagnosis were excluded from the analysis. Cancer stages has four categories: 'local', 'regional', 'distant', and 'unstaged' (not enough information to assign a stage). We excluded the unstaged category, and combined local and regional, leading to a dichotomized 'Stage' variable, as 'localized/regional (1)', or 'distant (0)'. Other patient-specific information used in our analysis are Age (at diagnosis, in complete years), Marital status (at

diagnosis; single/married/other), survival/censoring time after diagnosis and the (right)-censoring indicator. The survival times in the database represent total number of months the patient survived. To avoid zero survival time, we recoded the original times 0–71 to 1–72. All categorical variables (Race, Stage, and Marital Status) were dummy-coded. Spatial effects corresponding to the 64 counties were accommodated in the model by introducing 63 (spatial) frailty terms (recall that the first frailty term is set to be always 0), and the intercept term is represented by the baseline hazard.

Although we were successful in deriving the necessary Gibbs and Metropolis-Hastings steps for our MCMC implementation, the computational complexity was substantial. We used 2 chains with widely dispersed initial values. The first 5000 iterations were discarded as burn-in, and posterior summaries were computed using 5000 more iterations. MCMC convergence was monitored using trace plots, auto-correlation function plots, and Gelman-Rubin potential scale-reduction factor $\hat{R}$ (Gelman et. al, 2014). For the PrCA data with $n = 40907$ subjects, 1400 seconds of CPU time were required on a laptop with Intel(R) Core(TM) i5-3337U CPU 1.80 GHz. The codes are available from the <m>GitHub</m> link of the first author at: <https://github.com/achernoukhov/Bayesian-Spatial>.

4.2. Choice of the number of time intervals

Our choice of the optimal number of intervals (or break-points) is based on popular Bayesian model choice criteria, such as the Deviance Information Criterion, or (DIC) (Spiegelhalter, 2002), and measures based on posterior predictive performance (Gelman et. al, 2014). The DIC is defined as $DIC = \overline{D(\Omega)} + \tau_D$, where Ω is the full parameter space,

$\overline{D(\Omega)} = -2E\{\log[f(\mathbf{y}|\Omega)]|\mathbf{y}\}$, $f(\mathbf{y}|\Omega) = \prod_{i=1}^n f(y_i|\Omega)$ is the likelihood function, $E\{\log[f(\mathbf{y}|\Omega)]|\mathbf{y}\}$ is the posterior expectation of $\log[f(\mathbf{y}|\Omega)]$ and τ_D is a measure of the effective number of parameters in the model, given by $\tau_D = \overline{D(\Omega)} - D(\overline{\Omega})$, where $D(\overline{\Omega})$ is the deviance

computed at the posterior means of the parameters. Model selection follows the ‘lower is better’ rule, i.e., the model with the lowest DIC is selected. We also use the conditional predictive ordinate (CPO) statistic (Gelfand et al., 1992) for our model selection, derived from the posterior predictive distribution (ppd). Let $\mathcal{D}$ be the full data, $\mathcal{D}^{(-i)}$ the data with the i th observation deleted, and Ω , our parameter vector. We denote the posterior density of Ω , given $\mathcal{D}^{(-i)}$ by $\pi(\Omega|\mathcal{D}^{(-i)})$. For the i -th observation, the CPO_i can be written as

$CPO_i = \int f(y_i|\Omega)\pi(\Omega|\mathcal{D}^{(-i)})d\Omega = \left\{ \int (\pi(\Omega|\mathcal{D})/f(y_i|\Omega))d\Omega \right\}^{-1}$. In absence of a closed form, a Monte Carlo estimate of CPO_i can be obtained using a harmonic-mean approximation

(Dey et al., 1997) as $\widehat{CPO}_i = Q \left\{ \sum_{q=1}^Q (1/f(y_i|\Omega_q)) \right\}^{-1}$, where $\Omega_1, \dots, \Omega_Q$ is a post burn-in sample of size Q from $\pi(\Omega|\mathcal{D})$. A summary statistic of the CPO_i 's is the log pseudo-

marginal likelihood (LPML), defined by $LPML = \sum_{i=1}^n \log(\widehat{CPO}_i)$. Larger values of LPML indicates better fit.

The values of DIC , p_D and $LPML$ for various choices of the number of breakpoints m are presented in Table 1. One can notice that the effective number of parameters is much less than the actual number of parameters in the model. Also, with increase in m , it increases even slower than the actual number of parameters. For example, the effective number of parameters changes from 953 ($m = 36$) to 1,273 ($m = 72$), while the actual number of parameters increases from 2,520 to 5,040, implying that the penalty for introducing more parameters in the model is very small, and simultaneously, the fit of the model significantly improves when we use more time breakpoints, as indicated by the rapid decrease of DIC and increase of $LPML$. However, further increase of number of breakpoints (from $m = 72$) is unreasonable, given that the survival is presented in integer number of months. Therefore, the smallest reasonable partitioning is of length 1, leading to the maximum number of intervals as 72. From the Table, $m = 72$ gives us the smallest DIC and largest $LPML$. Hence, we select this model as our model of choice for subsequent data analysis.

4.3. Results

The estimates of the time-varying effects of the covariates are plotted in Figure 1. As far as the estimates of time the varying frailties concerned, we found that only four counties seemed to have no-zero frailties, and only those four frailties are reported in Figure 2. In each of these plots, the solid line indicates the (posterior) median of the parameters and frailties, while the dashed lines represent the corresponding 2.5% and 97.5% quantiles. We should remember here that County 1 (Acadia Parish) was retained as a reference county and hence its frailty was forced to 0.

In Figure 1, the regression functions for the effects of Age, Race, Stage, Marital Status (Married versus Other), and Marital Status (Single versus Other) are reported. Also, reported in Figure 2 is the the baseline hazard corresponding to a person with 'Age=0', 'Race=Black', 'Stage=Distant', and 'Marital Status=Married' from the Acadia Parish. From Figure 1 (first panel), the hazard for such person remains small before in the first month, increases rapidly between 1–2 months, and then decreases until it reaches its initial value at 5 months. Following this, it slowly decreases until 30 months, and remains approximately constant until 50 months, after which we again observe an increasing trend. The credible intervals (CIs henceforth) for the baseline hazard starting from 65 months are very wide, implying estimation uncertainties in this time period. On the contrary, the credible limits for the first 4 months are almost equal to the median value, implying a high degree of accuracy in estimation of the baseline hazard. Such a difference in accuracy is explained by the number of events (deaths) at these regions. The total number of events in the region $t \leq 65$ is 57 with only 2 events corresponding to the interval 71–72 months and a little more for the preceding intervals. For the interval 0–1 months, on the other hand, there are 2091 events which allows to estimate the hazard with a very good accuracy. The Age effect (from Regression function 1) appear significant during the first 5 months, after which it is no longer significant, given that the CIs always contain zero. Effect of Race (from regression function 2) also appears insignificant, however, the CIs are skewed towards negative values, which is indicative of a tendency for the black persons to have greater hazard than white. The Stage effect vary significant during the first 4 months, after which its effect diminishes. From the plots of regression functions 4 and 5, Marital Status is no longer significant after

ten months, within which the hazards for both Single and Other categories are greater than Married, and the hazard for the Other category is slightly greater than the Single. While the effect after 10 months dissipates, hazards for both Single and Other categories seem to remain one notch higher than the Married.

Most of the estimated spatial frailties appeared insignificant. However, the estimates from the 36th (Orleans) and 52nd (Saint Tammany) counties tend to be negative, implying that the hazard of patients in these counties is less than that of patients from Acadia parish (the baseline county). For the 33rd (Madison parish) and 54th (Tensas parish), however, the frailties are always positive which indicates larger hazards in these counties. These results are in agreement with those presented in Zhang and Lawson (2011). However, note that the AFT model used by Zhang and Lawson (2011) regresses covariates and frailties on the survival time, while our model considers the effects on the hazard through our AH formulation. So, the positive effect obtained in our model corresponds to the negative effect in theirs, and vice versa. Furthermore, our model can capture significant time-dependency of the estimated parameters and frailties, which, otherwise, cannot be caught by the AFT model.

5. CONCLUSIONS

In this paper, we developed a Bayesian additive hazards model with conditional autoregressive spatial frailties (AH-CAR) for the analysis of geo-referenced and right-censored survival data. Our proposed AH-CAR model extends the popular Aalen's AH model, and attempts to offer an alternative to the restrictive Cox's PH regression model by allowing time-varying regression coefficients as well as time varying spatial frailties. For estimation, we employed an MCMC approach that requires sampling from the conditionals of the parameters one at a time, given all other model parameters and the observed data.

Our motivating dataset was previously analyzed by Zhang and Lawson (2011) via an AFT model with CAR spatial frailties. In general, the AFT is designed to make direct inference about survival times, and hence, the effects of covariates are interpreted as the factors by which the median survival times are shortened or prolonged when comparing various levels of a given covariate. On the other hand, models based on the hazard function such the Cox's PH and the Aalen's AH model do not make direct inference about the survival times, but rather deal with the risk of developing the event of interest (e.g., death). Therefore, the hazard models and the AFT models are not directly comparable. Nevertheless, we posit our AH-CAR model to have a number of advantages over the AFT framework of Zhang and Lawson (2011). First, the AFT model is purely parametric relying heavily on the correct specification of the error distribution, while our AH-CAR is a semi-parametric model which does not make any distributional assumption on the survival time. Secondly, their AFT model does not accommodate time-varying effects (neither in regression coefficients nor in frailties), and therefore may even miss exposures whose effects change over time. This is revealed from the data analysis in Section 3, where our results agreed with theirs (effects of age and race were constant over time with similar magnitude), while the effect of stage and marital status in our model appeared to be time-varying. Finally, the coefficients of the AH

model are easily interpretable as absolute change in hazard for a unit change in exposure, while such interpretation is not readily available in the AFT model.

Although the current development is designed exclusively for the right-censored setup, extensions of the AH-CAR framework to other kinds of censoring, such as interval-censored, current status data, etc. are feasible. Here, partitioning of time was achieved via piecewise constant functions. There remain possibilities of accommodating various other alternative semi-parametric propositions, such as the Gamma process on the baseline hazards (Ibrahim et al., 2004), etc, to attain further flexibility. In addition, our current application to lattice data (those that appear in disease mapping problems) can certainly be modified to the geostatistical setting (Banerjee and Carlin, 2004). All these remain viable options for future research, and will be considered elsewhere.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The first three authors are grateful to the National Science and Engineering Research Council of Canada (NSERC) for the discovery grants that supported this research. Bandyopadhyay's research was partially supported by grants R03DE023372 and R01DE024984 from the US National Institutes of Health.

References

- Aalen O. Lecture Notes in Statistics: Mathematical Statistics and Probability Theory. NY: Springer; 1980. A model for nonparametric regression analysis of counting processes; 1–25.
- Aalen O. Mathematical Statistics and Probability Theory. NY: Springer; 1980. A Model for Nonparametric Regression Analysis of Counting Processes.
- Aalen OO. A linear regression model for the analysis of life times. *Statistics in Medicine*. 1989; 8:907–925. [PubMed: 2678347]
- Banerjee S, Wall MM, Carlin BP. Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics*. 2004; 4:123–142.
- Banerjee S, Carlin BP, Gelfand AE. *Hierarchical Modeling and Analysis for Spatial Data*. NJ: CRC Press; 2004.
- Bellera BA, MacGrogan G, Debled M, de Lara CT, Brouste V, Mathoulin-Plissier S. Variables with time-varying effects and the Cox model: Some statistical concepts illustrated with a prognostic factor study in breast cancer. *BMC Medical Research Methodology*. 2010;10–20. [PubMed: 20128907]
- Banerjee S, Dey D. Semiparametric proportional odds models for spatially correlated survival data. *Lifetime Data Analysis*. 2005; 11:175–191. [PubMed: 15938545]
- Besag J, York J, Mollie A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*. 1991; 43:1–20.
- Besag J, Kooperberg C. On Conditional and Intrinsic Autoregression. *Biometrika*. 1995; 82:733–746.
- Cai J, Zeng D. Additive mixed effect model for clustered failure time data. *Biometrics*. 2011; 67:1340–1351. [PubMed: 21418052]
- Casella G, George EI. Explaining the Gibbs Sampler. *The American Statistician*. 1992; 46:167–174.
- Chernoukhova A. Thesis, University of Windsor, Canada. 2013. Bayesian Spatial Additive Hazard Model.
- Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1972; 34:187–220.

- Cressie N, Wikle CK. Statistics for Spatio-Temporal Data. NJ: Wiley; 2011.
- Dey DK, Chen MH, Chang H. Bayesian Approach for the Nonlinear Random Effects Models. *Biometrics*. 1997; 53:1239–1252.
- Gamerman D, Lopes HF. Markov chain Monte Carlo: Stochastic simulation for Bayesian inference. NY: CRC Press; 2006.
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. NY: Chapman & Hall/CRC; 2014.
- Gelfand AE, Smith FM. Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*. 1990; 85:398–409.
- Gelfand AE, Smith FM, Lee TM. Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling. *Journal of the American Statistical Association*. 1992; 87:523–532.
- Gelfand AE, Dey DK, Chang H. Model determination using predictive distribution with implementation via sampling based methods. UK: Oxford University Press; 1992.
- Hastings WK. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*. 1970; 57:97–109.
- Heloise G, Lacourt A, Leffondre K. On the proportional hazards model for occupational and environmental case-control analyses. *Bmc Medical Research Methodology*. 2013; 13(18)
- Hussein A, Nkurunziza S, Tomanelli K. Efficient estimation for Aalen's additive hazards model. *Australian and New Zealand journal of statistics*. 2013
- Ibrahim JG, Chen MH, Sinha D. Bayesian Survival Analysis. NY: Springer; 2004.
- Klein JP, Moeschberger ML. Survival analysis: Techniques for censored and truncated data. NY: Springer; 2006.
- Kneib T. Mixed model-based inference in geoadditive hazard regression for interval-censored survival times. *Computational Statistics & Data Analysis*. 2006; 51:777–792.
- Knox KL, Bajorska A, Feng A, Tang W, Wu P, Tu XM. Survival analysis for observational and clustered data: an application for assessing individual and environmental risk factors for suicide. *Shanghai Arch Psychiatry*. 2013; 25(3):183194.
- Lawson AB. Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology. NY: Chapman & Hall/CRC Interdisciplinary Statistics; 2013.
- Jie LiYili HongRam ThapaBurkhart Harold E. Survival Analysis of Lob lolly Pine Trees With Spatially Correlated Random Effects. *Journal Of The American Statistical Association*. 2015; 110(510): 486–502.
- Lin DY, Yingieni Z. Semiparametric analysis of the additive risk model. *Biometrika*. 1994; 81:61–71.
- Lin DY, Zhiliang Ying. Additive Hazards Regression Models for Survival Data. *Proceedings of the First Seattle Symposium in Biostatistic*. 1997:185–198.
- Lin PS. Analysis of spatial frailty models by a weighted estimating equation. *Journal of Statistical Planning and Inference*. 2012; 142:1436–1444.
- Lin FC, Zhu J. Additive hazards regression and partial likelihood estimation for ecological monitoring data across space. *Statistics and its Interface*. 2012; 5:195–206.
- Aaalen O, Borgan A, Gjessing H. Survival and Event History Analysis: A Process Point of View. NY: Springer; 2008.
- Huan J, Patrick BE, Havard R, Silvia S. Geostatistical survival models for environmental risk assessment with large retrospective cohorts. *Journal Of The Royal Statistical Society Series A-Statistics In Society*. 2014; 177(3):679–695.
- Powell LL, Wolfe Jared D, Johnson Erik I, Hines James E, Nichols James D, Stouffer Philip C. A Heterogeneous movement of insectivorous Amazonian birds through primary and secondary forest: A case study using multistate models with radiotelemetry data. *Biological Conservation*. 2015; 188(SI):100–108.
- Silva GL, Amaral-Turkman MA. Bayesian Analysis of an Additive Survival Model with Frailty. *Communications in Statistics – Theory and Methods*. 2004; 33:2517–2533.

- Surveillance, Epidemiology, and End Results (SEER) Program, Limited-Use Data. National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch. 2008. Released April 2008, based on the November 2007 submission, 1973–2005, www.seer.cancer.gov
- Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2002; 64:583–639.
- Wolfson LJ, Kadane JB, Small MJ. Bayesian Environmental Policy Decisions: Two Case Studies. *Ecological Applications*. 1996; 6(4):1056–1066.
- Zhang J, Lawson AB. Bayesian parametric accelerated failure time spatial model and its application to prostate cancer. *Journal of Applied Statistics*. 2011; 38:591–603. [PubMed: 21475617]

Figure 1.
SEER Data: Plots of time-varying regression coefficients

Figure 2.
Plots of estimated time-varying frailties corresponding to Louisiana counties 33, 36, 52 and 54

Table 1

Model comparison: Values of DIC , p_D and $LPML$ for varying number of intervals m .

Number of Breakpoints m	Actual Number of Parameters n_{par}	Effective Number of Parameters p_D	Deviance Information Criterion DIC	Conditional Predictive Ordinate (aggregated) $LPML$
1	70	63	189,244	-94,630
12	840	463	-692,284	327,009
24	1,680	740	-1,780,091	827,536
36	2,520	953	-3,098,172	1,403,102
72	5,040	1,273	-7,702,954	3,393,691