# Bayesian Spatiotemporal Modelling of Survival Outcomes in Long-Term Studies

**Ziyu Zheng, Barry Rowlingson and Benjamin Taylor**

CHICAS, Lancaster Medical School,
Lancaster University

z.zheng@lancaster.ac.uk

## Introduction

Population-based cancer registries usually collect data over an extended period of time e.g. decades. Individuals enter these registries when they are diagnosed with a cancer and their survival prognoses will depend on the effectiveness of available treatment regimens at the time of entry and also the rapidity with which the cancer was detected. More generally, improvements in survival prognosis can be seen not only as a result of treatment quality or public health awareness but also as a result of biases associated with screening: over-diagnosis, lead time and length biases.

In many studies, interest focuses on survival duration and the time of entry into the study is eliminated. In this poster, we introduce a Bayesian spatiotemporal model that captures both entry-time effects as they evolve over the study period and also individual survival duration. We use a transformed Gaussian process to model the time of entry as a risk factor for survival as well as the time interval after enrolment.

## Motivational Example

**Example 0.1.** This example demonstrates how we want to take into account of both survival and entering times. Consider a study of 10 individuals with different entering and survival times as shown in Figure 1.

- Time is partitioned into 4 intervals; they do NOT have to have equal length.
- We aim to model both the duration times and the entering times for individuals.
  - Duration times are dealt with based on ordinary survival analyses.
  - Entering times are captured by introducing some stationary temporal Gaussian processes.

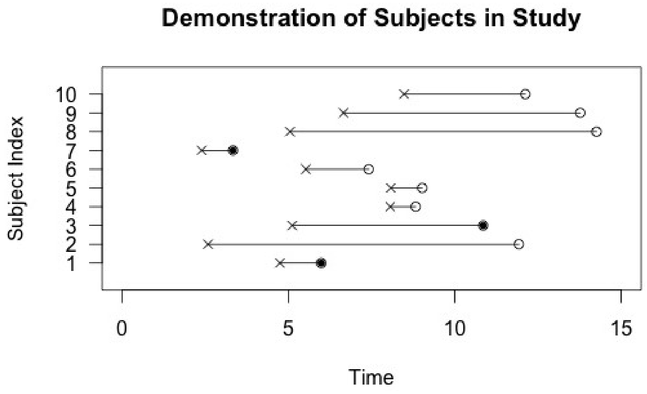


**Figure 1:** Example: Time of Entry and Censoring for Individuals

## Methods

### Ornstein-Uhlenbeck process [1]

Suppose that $\{Y_t\}$ is a stationary temporal Gaussian process with a separable covariance function such that

$$Y_t = a_{t-1,t}Y_{t-1} + [1 - a_{t-1,t}]\mu + [1 - a^2_{t-1,t}]^{1/2}\epsilon_t, \tag{1}$$

where $a_{t-1,t} \in [0, 1]$ and $\epsilon_t \sim \mathcal{N}(0, \Sigma_Y)$.
Given $Y_{t-1}$, one can simulate $Y_t$ via:

$$\begin{aligned} Y_t =& a_{t-1,t}Y_{t-1} + [1 - a_{t-1,t}]\mu + [1 - a^2_{t-1,t}]^{1/2}\Sigma_Y^{1/2}\mathcal{N}(0, 1) \\ =& a_{t-1,t}(\mu + \Sigma_Y^{1/2}\gamma_{t-1}) + (1 - a_{t-1,t})\mu + \sqrt{1 - a^2_{t-1,t}}\Sigma_Y^{1/2}\mathcal{N}(0, 1) \\ =& \mu + a_{t-1,t}\Sigma_Y^{1/2}\gamma_{t-1} + \sqrt{1 - a^2_{t-1,t}}\Sigma_Y^{1/2}\mathcal{N}(0, 1) \end{aligned} \tag{2}$$

### Temporal Baseline Hazard Function

Recall the time intervals shown in Figure 1 $I_j = (I_j^{\text{lower}}, I_j^{\text{upper}})$, then for $t + t_0$ in $I_j$, we define $a_{t-1,t} = \exp\{-\theta\Delta t\}$, where $\Delta t = (t + t_0) - I_j^{\text{lower}}$. The new proposed adjustment of the model should allow both duration time (survival time $t$) and the real time scale (captured as entering time $t_0$ take roles in the model.

$$h_0 = f(t; \omega_f) \exp\{g(t + t_0; \tau, \theta, \gamma^{(t)})\}$$

$f(t; \omega_f)$ : some ordinary baseline hazard function, e.g. Weibull hazards;

$g(t + t_0; \tau, \theta, \gamma^{(t)})$ : some stationary temporal Gaussian process. More precisely, we define

$$g(t + t_0; \tau, \theta, \gamma^{(t)}) = \sum_{j=1}^{m}(\tau\gamma_j^{(t)} - \frac{\tau^2}{2})\mathbb{I}(t + t_0 \in I_j). \tag{3}$$

**Note** : more details about $g$ can be found in later sections.

**Example 0.2.** Figure 2 shows an example of the proposed baseline hazard function for an individual in the study. Here:

- Time is partitioned into 4 intervals by the quantile method;
- The intervals are inclusive of the lower ends
- $f$ is some Weibull hazard function with $\alpha = 0.42$ and $\lambda = 2.57441$.
- In this example, the individual entered study in $I_1$ and the value $t + t_0$ falls in the last time interval $I_4$, thus

$$g(t + t_0; \tau, \theta, \gamma^{(t)}) = \tau\gamma_4^{(t)} - \frac{\tau^2}{2}.$$

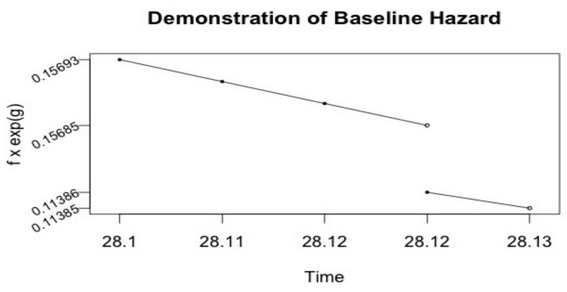


**Figure 2:** Example: Plot of Proposed Temporal Baseline Hazard

### Temporal Cumulative Hazard Function

Define $j^*_{\text{max}} = \max_j\{t + t_0 \in I_j\}$ and $j^*_{\text{min}} = \min_j\{t_0 \in I_j\}$, denote $g^{(j)}$ to be the value of function $g$ when $t + t_0 \in I_j$, the cumulative baseline hazard thus follows:

$$\begin{aligned} H_0(t) =& \int_0^t f(s)\exp\{g(s + t_0)\}\mathrm{d}s \\ =& \sum_{j^*_{\text{min}}+1}^{j^*_{\text{max}}-1}\exp\{g^{(j)}\}\int_{I_j} f(s)\mathrm{d}s + \exp\{g^{(j^*_{\text{min}})}\}\int_{t_0}^{I^{\text{upper}}_{j^*_{\text{min}}}} f(s)\mathrm{d}s + \exp\{g^{(j^*_{\text{max}})}\}\int_{I^{\text{lower}}_{j^*_{\text{max}}}}^{t} f(s)\mathrm{d}s. \end{aligned}$$

### Spatio-Temporal Hazard Function

Based on the proposed temporal baseline hazard function $h_0$, the hazard function follows:

$$h(t_i; \phi, Z_i) = \exp\{X_i\beta + Z_i\}h_0(t_i; \omega),$$

$Z$ : some spatially continuous stationary latent Gaussian field, where $Z_i$ is the value at location of observation $i$;

$\omega$ : vector of parameters in $h_0$; eg. $\omega = (\omega_f, \tau, \theta, \gamma^{(t)})$;

$\phi = (\beta, \omega, \eta)$ : parameters of covariate effects, baseline hazard and parameters of the covariance function of $Z$ respectively.

The Exponential model can be one suitable proposal for Cov$(Z)$; ie. $\sigma^2\exp\{-d/\Phi\}$ for $\sigma^2$ being the marginal variance of fields and $\Phi$ is the 'spatial decay' parameter.

### Inference

During the MCMC scheme, we will NOT sample the $Z$'s nor $Y$'s directly, rather with a vector of transformed variables, $\gamma = (\gamma_1^{(s)}, \ldots, \gamma_m^{(s)}, \gamma_1^{(t)}, \ldots, \gamma_n^{(t)})$.

**transformation of $Y$'s**

- Recall Equation 3, for each $I_j$ we define the relationship between $\gamma_j^{(t)}$ and $Y_t$ to be $\gamma_j^{(t)} = \tau^{(-1/2)}(Y_t - \mu)$, where $Y_0 \sim \mathcal{N}(\mu, \tau^2)$ implies $Y_t \sim \mathcal{N}(\mu, \tau^2)\ \forall t$.
- $\gamma_t^{(t)}|\gamma_{t-1}^{(t)} \sim \mathcal{N}(a_{t-1,t}\gamma_{t-1}^{(t)}, 1 - a^2_{t-1,t})$ as the updating procedure in Equation 2 follows after substituting in $\gamma_t$, :

$$\begin{aligned} \gamma_t^{(t)} =& \tau^{-1}[\mu + a_{t-1,t}\tau\gamma_{t-1}^{(t)} + \sqrt{1 - a^2_{t-1,t}}\tau\mathcal{N}(0, 1) - \mu] \\ =& a_{t-1,t}\gamma_{t-1}^{(t)} + \sqrt{1 - a^2_{t-1,t}}\mathcal{N}(0, 1) \end{aligned}$$

- The latent process $\{Y_t\}$ is parameterised such that $\mathbb{E}[\exp(Y)] = 1$ by setting $\mu$ to be $-\tau^2/2$ for $\tau^2$ the marginal variance. Apriori, $\gamma_0^{(t)} \sim \mathcal{N}(0, 1)$, $\gamma^{(t)}$ can then be evaluated from the decomposition, $\forall T$:

$$\pi(\gamma_0^{(t)}, \ldots, \gamma_T^{(t)}) = \pi(\gamma_0^{(t)})\pi(\gamma_1^{(t)}|\gamma_0^{(t)}) \ldots \pi(\gamma_T^{(t)}|\gamma_{T-1}^{(t)}).$$

- Inference proceed in a similar fashion to [2] by sampling from the joint density $\pi(Y, Z, \phi|\text{data})$.
- Note: The transformation of $Z$ takes a very similar path; referred to as $\gamma^{(s)}$.

**MCMC**

Inference method is an example of a Metropolis-Hastings sampling scheme. Samples are drawn from the posterior $\pi(\phi, \gamma^{(s)}|\text{data}) \propto \pi(\text{data}|\phi, \gamma^{(s)})\pi(\phi, \gamma^{(s)})$, using MCMC [3, 4] where the parameters are transformed; eg. $\tilde{\omega} = \log\omega$. The density $\pi(\text{data}|\phi, \gamma^{(s)}) = \pi(\text{data}|\beta, \tilde{\omega}, \gamma^{(s)})$ due to the conditional independence.

- The MCMC scheme has Langevin kernels for $\beta$, $\tilde{\omega}_f$, $\gamma^{(s)}$, $\tilde{\tau}$, $\tilde{\theta}$, $\gamma^{(t)}$ and a random walk kernel for $\tilde{\eta}$.
- Algorithm is an example of Metropolis-Hastings sampling:
  - Initialise the chain at $\{\beta^{(0)}, \tilde{\omega}^{(0)}, \tilde{\eta}^{(0)}, \gamma^{(s)}_{(0)}\}$;
  - Proposal density for $\zeta = (\beta, \tilde{\omega}, \tilde{\eta}, \gamma^{(s)})$ is $q(\zeta(i^*)|\zeta^{(i-1)}) = \mathcal{N}(\zeta^{(i^*)}; \mu_{\zeta^{(i-1)}}, h^2\Sigma)$, where

$$\mu_{\zeta^{(i-1)}} = \begin{bmatrix} (\beta, \tilde{\omega})^{(i-1)} + \frac{h^2h^2_{\beta,\tilde{\omega}}}{2}\Sigma_{\beta,\tilde{\omega}}\frac{\partial\log\{\pi(\zeta^{(i-1)}|Y)\}}{\partial(\beta,\tilde{\omega})} \\ \tilde{\eta}^{(i-1)} \\ \gamma^{(s)}_{(i-1)} + \frac{h^2h^2_{\gamma^{(s)}}}{2}\Sigma_{\gamma^{(s)}}\frac{\partial\log\{\pi(\zeta^{(i-1)}|Y)\}}{\partial\gamma^{(s)}} \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} h^2_{\beta,\tilde{\omega}}\Sigma_{\beta,\tilde{\omega}} & 0 & 0 \\ 0 & ch^2_{\tilde{\eta}}\Sigma_{\tilde{\eta}} & 0 \\ 0 & 0 & h^{2(s)}_{\gamma}\Sigma_{\gamma^{(s)}} \end{bmatrix}.$$

  - Here the constants $h^2_{\cdot}$ are optimal scalings in MALA proposals [6]; More details can be seen in R package spatsurv [7]. The optimal value of $h$ should give asymptotic acceptance rate 0.574.

## Test Results

This method is tested on the skin cancer data in California between $1973 - 2013$ [5]. Here the test model considered age and sex as covariates. In the baseline hazard function, $t_0$ and $t$ are the diagnostic and survival times respectively. For testing purpose, we only worked with 1000 random samples from the whole dataset.

The test model ran 500000 iterations with burning length 1000 and thining 50. The MCMC chain plot has shown satisfactory convergence and mixing with scaling parameter $h$ being close to 1.

- Fixed effects and parameters for baseline hazard and spatial covariance are shown as below:

| | Sex | Age | $\alpha$ | $\lambda$ | $\tau$ | $\theta$ | $\gamma_1^{(t)}$ | $\gamma_2^{(t)}$ | $\gamma_3^{(t)}$ | $\gamma_4^{(t)}$ | $\sigma$ | $\phi$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **mean** | **0.85** | **1.05** | **0.63** | **0.04** | **2.90** | $\mathbf{6.94 \times 10^{-3}}$ | **0.28** | **0.23** | **0.32** | **0.52** | **0.71** | $\mathbf{1.10 \times 10^5}$ |
| 2.5% | 0.67 | 1.04 | 0.57 | 0.02 | 2.64 | $4.66 \times 10^{-3}$ | 0.10 | 0.05 | 0.14 | 0.24 | 0.53 | $7.30 \times 10^4$ |
| 97.5% | 1.12 | 1.06 | 0.70 | 0.08 | 3.16 | $1.02 \times 10^{-2}$ | 0.48 | 0.43 | 0.52 | 0.74 | 0.99 | $1.65 \times 10^5$ |

**Table 1:** Baseline Hazard Parameters and 95% Confidence Bands

- The influence of the choice of prior (log-scale prior in red) on the resulting parameter estimates (histogram); example is shown in Figure 3. For each parameter, a big difference between the prior and posterior, indicates that the data are informative about that parameter. Figure 4 shows the posterior median of the spatial covariance function and 95% credible interval;
- Figure 5 shows the posterior probability that the covariate-adjusted relative risk greater than 1.1. Areas of high probability (in red) in this plot are where the relative risk of death occurring adjusted for age and sex, is such that there may be cause for concern by public health authorities;
- Figure 6 shows the values of $\exp(g)$ over time with the 95% credible interval.

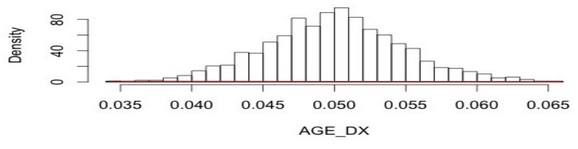


**Figure 3:** Prior (red line) and posterior (histogram) for age

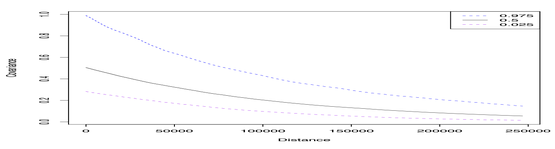


**Figure 4:** Posterior median of the spatial covariance

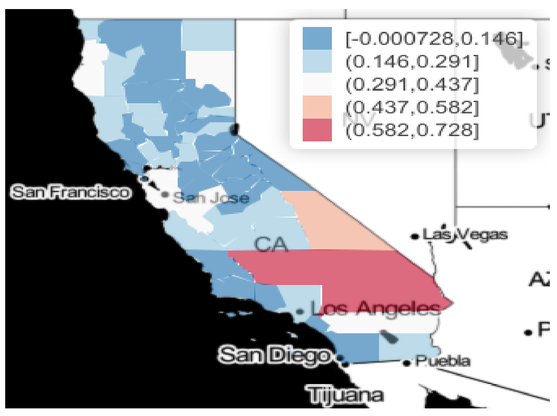


**Figure 5:** Map of Relative Risk

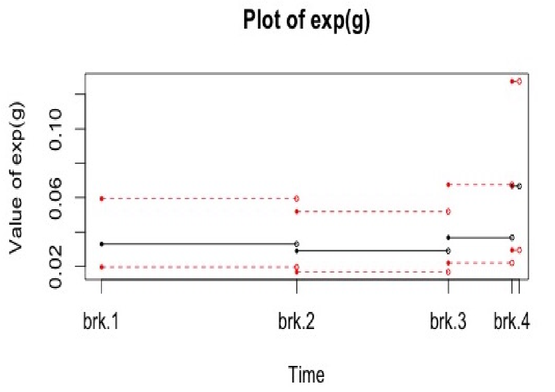


**Figure 6:** Plot of $\exp\{g\}$ over real time; brk denote break points for intervals.

## Forthcoming Research

We will be working with the full cancer dataset in the future; the method shall be further developed to include but not limited to the followings:

- the current baseline hazard function should be extended to allow free survival and entering times for evaluation of hazard functions and time varying covariates, say $\omega(t + t_0)$;
- censoring types can be extended to include interval censoring;
- developing methods for spatio-temporal prediction over data;
- reducing the computational costs.

## References

[1] E. Bibbona, G. Panful, and P. Tavella. The ornstein-uhlenbeck process as a model of a low pass filtered white noise. *Metrologia*, 45, 2008.

[2] A. Brix and P. J. Diggle. Spatiotemporal prediction for log-gaussian cox processes. *Journal of the Royal Statistical Society, Series B.*, 63:823–841, 2001.

[3] D. Gamerman and H. F. Lopes. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference (2nd ed.)*. Chapman and Hall/CRC, 2006.

[4] Richardson S. Gilks, W. and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, 1995.

[5] N. Howlader, A. M. Noone, M. Krapcho, J. Garshell, N. Neyman, S. F. Altekrues, C.L. Kosary, M. Yu, J. Ruhl, Z. Tatalovich, H. Cho, A. Mariotto, D. R. Lewis, H. S. Chen, E. J. Feuer, and K. A. Cronin. *SEER Cancer Statistics Review 1975-2010*. National Cancer Institute., 2013.

[6] G. Roberts and J. Rosenthal. Optimal scaling for various metropolis-hastings algorithms. *Statistical Science*, 16:351–367, 2001.

[7] B. M. Taylor and B. S. Rowlingson. spatsurv: an R package for Bayesian inference with spatial survival models., 2014. Available from http://cran.r-project.org/web/packages/spatsurv/index.html.