A Spatial Copula Approach to Fully Bayesian Nonparametric Survival Analysis

UNIVERSITY OF SOUTH CAROLINA

Haiming Zhou, Timothy Hanson, and Roland Knapp

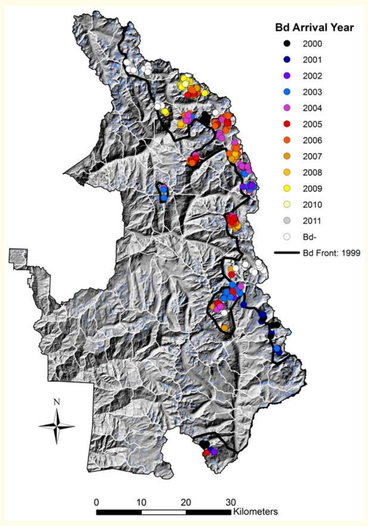
Background

► Frogs and other amphibians have been dying off in large numbers since the 1980s because of a deadly fungus called Batrachochytrium dendrobatidis, also known as Bd.

► Dr. Roland Knapp has been studying the amphibian declines for the past decade at Sierra Nevada Aquatic Research Laboratory.

► He hiked thousands of miles and surveyed hundreds of frog populations in Sequoia-Kings Canyon National Park.

Frog Data (2000-2011)



► The response is the Bd infection time (i.e. Bd arrival year − baseline year).

► Main covariates:
bdwater: whether or not Bd has been found in the watershed.
bddistance: straight-line distance to the nearest Bd location.

► It contains 309 frog populations. Each was followed up until infection or being censored (10% censoring).

► Spatial dependence: populations near each other tend to become infected at about the same time.

Objective

► $T(\boldsymbol{s})$: the random Bd infection time (i.e. survival time) at location $\boldsymbol{s}$.

► $\{T(\boldsymbol{s})|\boldsymbol{s} \in \mathcal{D}\}, \mathcal{D} \subseteq \mathbb{R}^2$: a spatial process.

► $(T(\boldsymbol{s}_1), \ldots, T(\boldsymbol{s}_n))'$: a realization.

► $\boldsymbol{x}(\boldsymbol{s})$: a $p \times 1$ vector of covariates.

► Goal: describe the association between $\boldsymbol{x}(\boldsymbol{s})$ and $T(\boldsymbol{s})$ while allowing for spatial dependence and predict $T(\boldsymbol{s}_0)$ given $\boldsymbol{x}(\boldsymbol{s}_0)$ at any new location $\boldsymbol{s}_0$.

Proportional Hazard Model

Assume $T_i|\boldsymbol{x}_i, i = 1, \ldots, n$ **independently** distributed with hazard rate

$$\lambda(t|\boldsymbol{x}_i) = \lambda_0(t)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}),$$

where $\lambda_0(t)$ is the baseline hazard of an individual with $\boldsymbol{x} = \boldsymbol{0}$.

Li and Lin (2006), JASA

Normal Transformation Model:

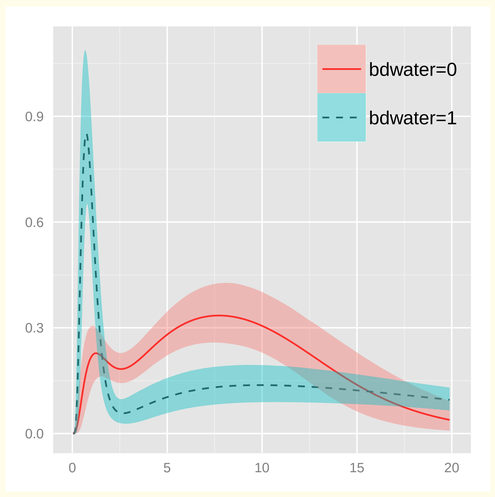
► Assume $T_i|\boldsymbol{x}_i$ follows the proportional hazard model with cdf $F_{\boldsymbol{x}_i}, i = 1, \ldots, n$.

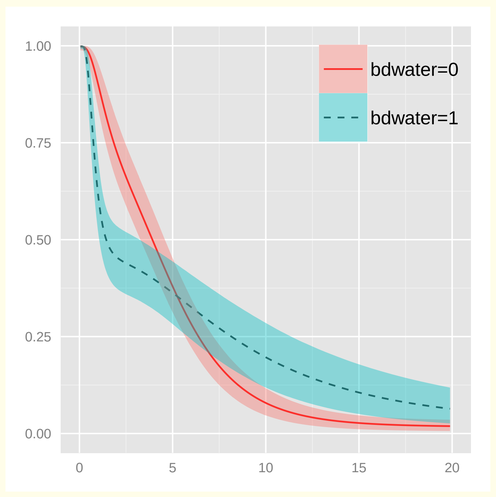
► Then $T_i^* = \Phi^{-1}\{F_{\boldsymbol{x}_i}(T_i)\} \sim N(0,1)$, where $\Phi(\cdot)$ is the cdf of the standard normal variable.

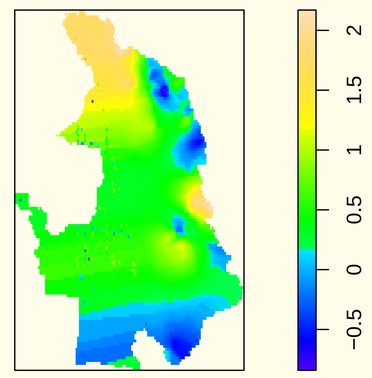
► Assume $\boldsymbol{T}^* = (T_1^*, \ldots, T_n^*) \sim N_n(\boldsymbol{0}, \boldsymbol{C})$, where $\boldsymbol{C}$ is a covariance matrix defined by the Matern correlation function.

Frog Data Analysis

	Our Approach (LMPL $= -277$)				Li&Lin (LMPL $= -326$)			
Parameters	Mean	Median	Std. dev.	95% HPD Interval	Mean	Median	Std. dev.	95% HPD Interval
θ_1	**0.9937**	0.9941	0.0029	(0.9879, 0.9988)	0.9909	0.9913	0.0040	(0.9829, 0.9981)
θ_2	**0.0866**	0.0841	0.0211	(0.0493, 0.1297)	0.1217	0.1200	0.0204	(0.0842, 0.1638)





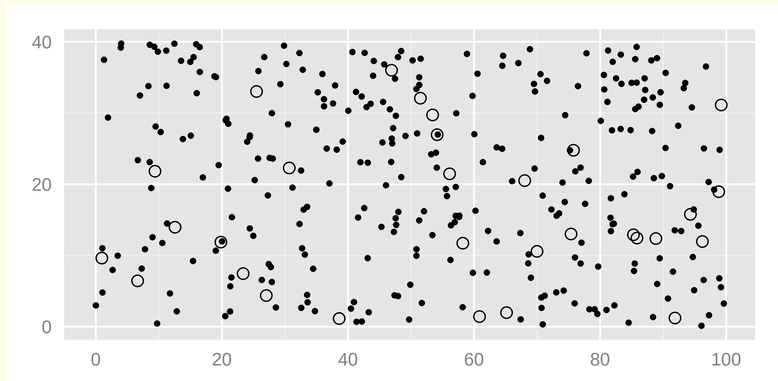


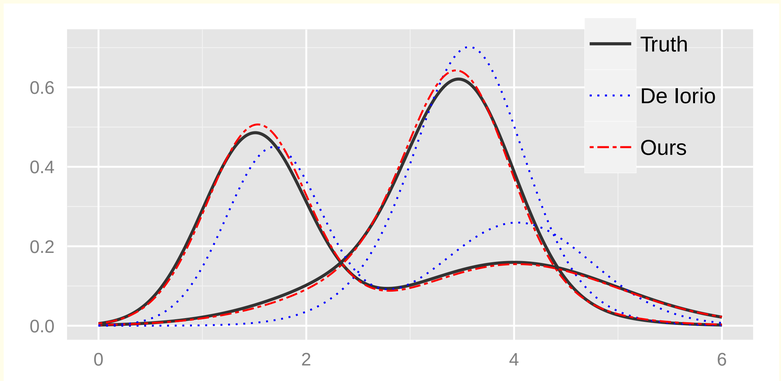
Observations:

1. Based on LPML, our approach has much better prediction ability than Li and Lin, 2006.
2. Based on the estimates of θ_1 and θ_2, the spatial correlation is strong for this data set.
3. The proportional hazard assumption does not hold since two curves are crossing.
4. The prediction map is very important to show which area will get infection more quickly.

Simulation

Assume $\log T|x$ follows $f(y|x) = 0.4N(3.5 + 0.5x, 1^2) + 0.6N(2.5 - x, 0.5^2)$, $\theta_1 = 0.98$ and $\theta_2 = 0.1$.





Mean Squared Prediction Error (MSPE): Ours $= \mathbf{0.269}$; Lin&Lin $= 0.498$; DeIorio $= 1.325$.

Nonproportional Hazard Model

De Iorio et al. (2009) assume $Y_i = \log T_i$ given $\boldsymbol{x}_i$ **independently** follows a mixture model

$$F_{\boldsymbol{x}_i}(y|G) = \int \Phi\left(\frac{y - \boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right) dG\{\boldsymbol{\beta}, \sigma^2\},$$

where G follows a Dirichlet Process (DP) prior.

A Spatial Copula Extension

► Assume $Y_i = \log T_i$ given $\boldsymbol{x}_i$ follows follows the above mixture model.

► Model the joint distribution of $(Y_1, \ldots, Y_n)'$ by

$$F(t_1, \ldots, t_n|G) = C(F_{\boldsymbol{x}_1}(t_1|G), \ldots, F_{\boldsymbol{x}_n}(t_n|G); \boldsymbol{\theta}),$$

where $C(u_1, \ldots, u_n; \boldsymbol{\theta})$ is a spatial copula with parameter $\boldsymbol{\theta}$, capturing spatial dependence .

► Define Gaussian copula as

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

where Φ_n be the joint cdf of $N_n(\boldsymbol{0}, \boldsymbol{C})$.

► Define spatial Gaussian copula by assuming $\boldsymbol{C} = [\boldsymbol{C}(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\theta})]_{i,j=1}^n$ with

$$\boldsymbol{C}(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\theta}) = \theta_1\rho(\boldsymbol{s}_i, \boldsymbol{s}_j) + (1 - \theta_1)I(\boldsymbol{s}_i = \boldsymbol{s}_j),$$

where $\theta_1 \in [0, 1]$ measures a local maximum correlation and $\rho(\boldsymbol{s}_i, \boldsymbol{s}_j) = \exp\{-\theta_2||\boldsymbol{s}_i - \boldsymbol{s}_j||\}$ is the correlation function.

MCMC

► All parameters involved in G are updated based on a modification of the blocked Gibbs sampler (Ishwaran and James, JASA, 2001): M-H samplers with independent proposals.

► Use Delayed Rejection (Tierney and Mira, 1999) if low acceptance rate occurs in M-H step.

► The correlation parameters $\boldsymbol{\theta}$ are updated using adaptive M-H (Haario et al., Bernoulli, 2001).

► For large n, the inversion of the $n \times n$ matrix $\boldsymbol{C}$ can be substantially speeded up using a full scale approximation (Sang and Huang, JRSSB, 2012).

A Future Direction

► Create a R package named "spBayesSurv".

► Consider other non-Gaussian spatial copulas.

► Goodness-of-fit testing for spatial copulas.

► Extend to categorial spatial data.

References

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

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