Joint Analysis of Multivariate Spatial Count and Zero-Heavy Count Outcomes

Using Common Spatial Factor Models

Cindy Feng
U Saskatchewan

Charmaine Dean
U Western Ontario

Outline

- Introduction
- Joint outcome analysis
- Testing for common spatial structure
- Precision gains through joint modeling
- Concluding remarks

Introduction

- Joint modelling
  - Account for dependence between longitudinal and time-to-event outcomes
- Scientific objectives for joint modelling
  - Inferences about longitudinal outcome, while accounting for informative drop-out time
  - Inferences about survival outcome, while accounting for association between the two outcomes
  - Interested in the relationship between the two outcomes

General model formulation

Longitudinal model $Y_i(t) = \mu_i(t) + W_{1i}(t) + \varepsilon_i(t)$

Survival model $\lambda_i(t) = \lambda_0(t) \exp\{ X_i\beta + W_{2i}(t) \}$

- $\mu_i(t)$ - mean response
- $\varepsilon_i(t)$ - measurement error
- X - covariate
- $W(t)=[W_{1i}(t)\ \ W_{2i}(t)]$ - bivariate latent Gaussian process

General model formulation

Longitudinal model $$Y_i(t) = \mu_i(t) + W_{1i}(t) + \varepsilon_i(t)$$

Survival model $$\lambda_i(t) = \lambda_0(t) \exp\{ X_i\beta + W_{2i}(t) \}$$

- Shared Frailties

$$W_{2i}(t) = \gamma W_{1i}(t) + V_i$$

Shared Frailties

- Used in a variety of contexts
- Dunson (2009), for example, for linking glmms

Common Spatial Factor Model for Count Data

Let $y_{ij}|\mu_{ij} \sim Poisson(\mu_{ij})$ for $i = 1, \cdots, n$ regions and $j = 1, \cdots, J$ outcomes. The model can be written as:

$$\log(\mu_{ij}) = \alpha_j + \log(E_{ij}) + \gamma_j b_i + h_{ij}$$

- $\mathbf{b} = (b_1, \cdots, b_n)' \sim \text{MVN}(\mathbf{0}, \sigma_b^2(\mathbf{D} - \mathbf{W})^{-1})$: spatially correlated random effects.
- γ_j: factor loading for the shared spatial component with $\gamma_1 \equiv 1$.
- $\mathbf{h}_j = (h_{1j}, \cdots, h_{nj})' \sim \text{MVN}(\mathbf{0}, \sigma_{h_j}^2 \mathbf{I})$: spatially uncorrelated random effects.

The joint posterior distribution is expressed as

$$p(\boldsymbol{\alpha}, \mathbf{b}, \mathbf{h}, \boldsymbol{\gamma}, \sigma_b^2, \sigma_{h_1}^2, \cdots, \sigma_{h_J}^2 | \mathbf{Y}) \propto L(\mathbf{Y}|\boldsymbol{\alpha}, \mathbf{b}, \mathbf{h}, \boldsymbol{\gamma}) p(\mathbf{b}|\sigma_b^2) p(\mathbf{h}|\sigma_h^2) \\ p(\boldsymbol{\alpha}) p(\boldsymbol{\gamma}) p(\sigma_b^2) p(\sigma_{h_1}^2) \cdots p(\sigma_{h_J}^2) .$$

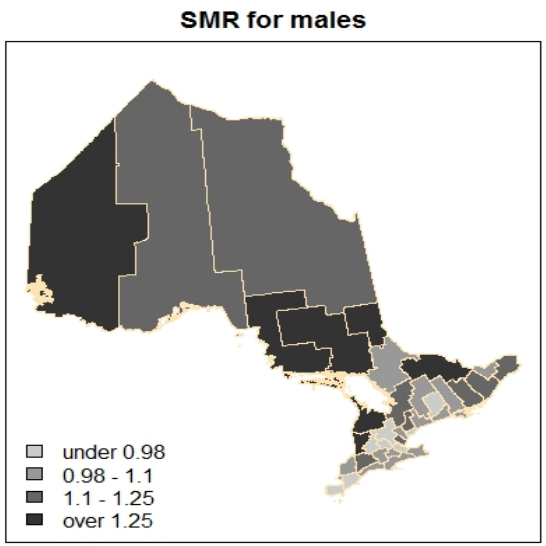
The first term is the conditional likelihood,

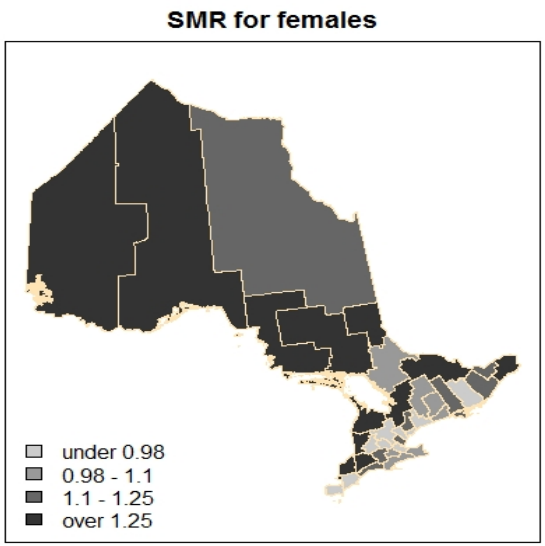
$$L(\mathbf{Y}|\boldsymbol{\alpha}, \mathbf{b}, \mathbf{h}, \boldsymbol{\gamma}) \propto \exp\left[- \sum_{i=1}^{n} \sum_{j=1}^{J} E_{ij} \exp(\alpha_j + \gamma_j b_i + h_{ij}) \right] \\ \times \prod_{i=1}^{n} \prod_{j=1}^{J} [E_{ij} \exp(\alpha_j + \gamma_j b_i + h_{ij})]^{y_{ij}} .$$

The second and third terms are the distributions of **b** and **h** respectively and the remaining terms are the prior distributions.

Ontario Lung Cancer

- Outcome: total observed counts of mortality from lung cancer for males and females over the period 1995 to 2002.
- Sub-regions: 37 local health areas (LHA) in Ontario.





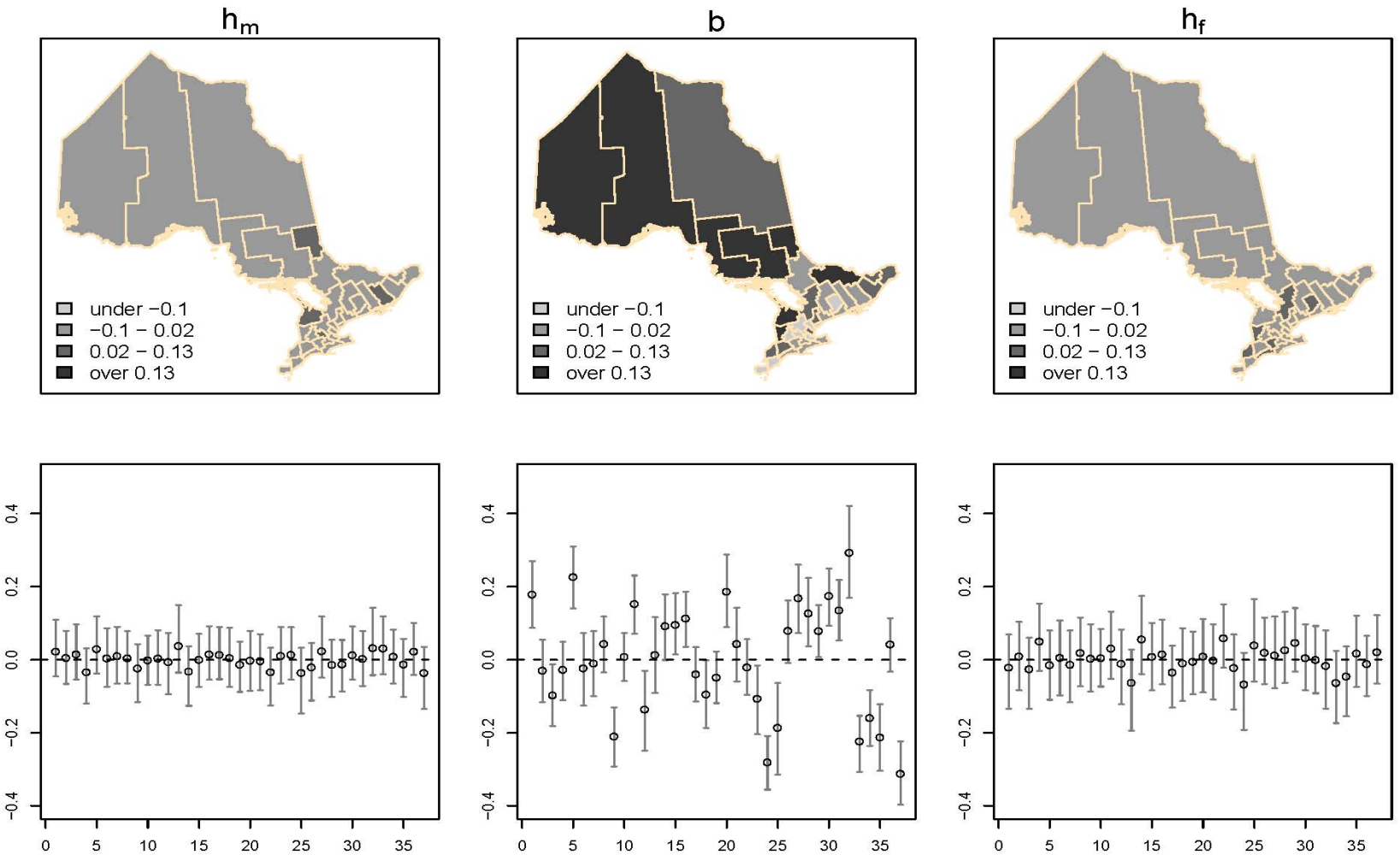
Let m and f index males and females, respectively. The common spatial factor model is,

$$\begin{cases} \log(\mu_{im}) = \alpha_m + \log(E_{im}) + b_i + h_{im} \\ \log(\mu_{if}) = \alpha_f + \log(E_{if}) + \gamma \cdot b_i + h_{if} \end{cases}$$

Table : Posterior Summaries

Parameter	Mean	95% CI
γ	1.145	(0.866, 1.466)
σ_b^2	0.059	(0.032, 0.116)
$\sigma_{h_m}^2$	0.0038	(0.0018, 0.0094)
$\sigma_{h_f}^2$	0.0050	(0.0023, 0.0129)

Ontario Lung Cancer



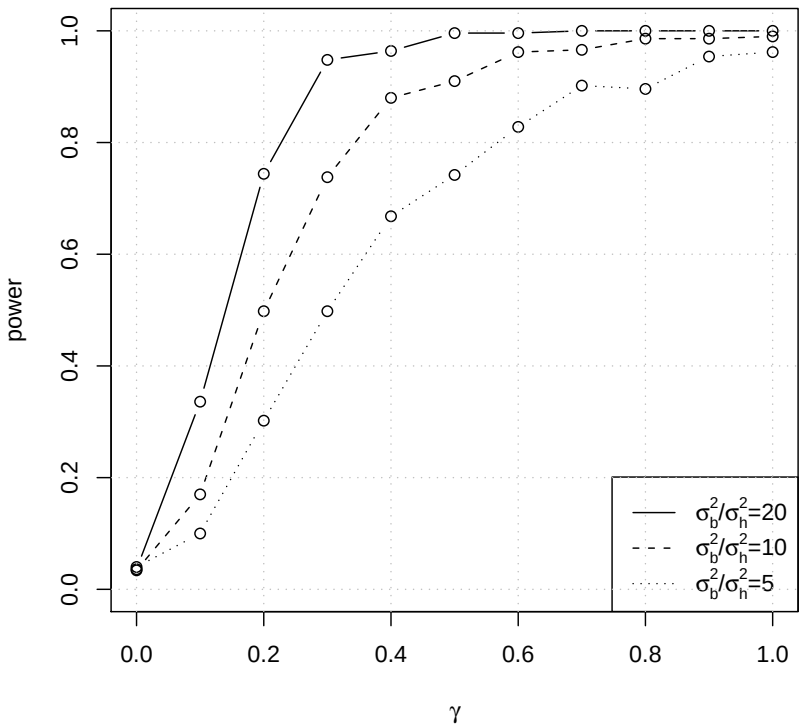
Ontario Lung Cancer

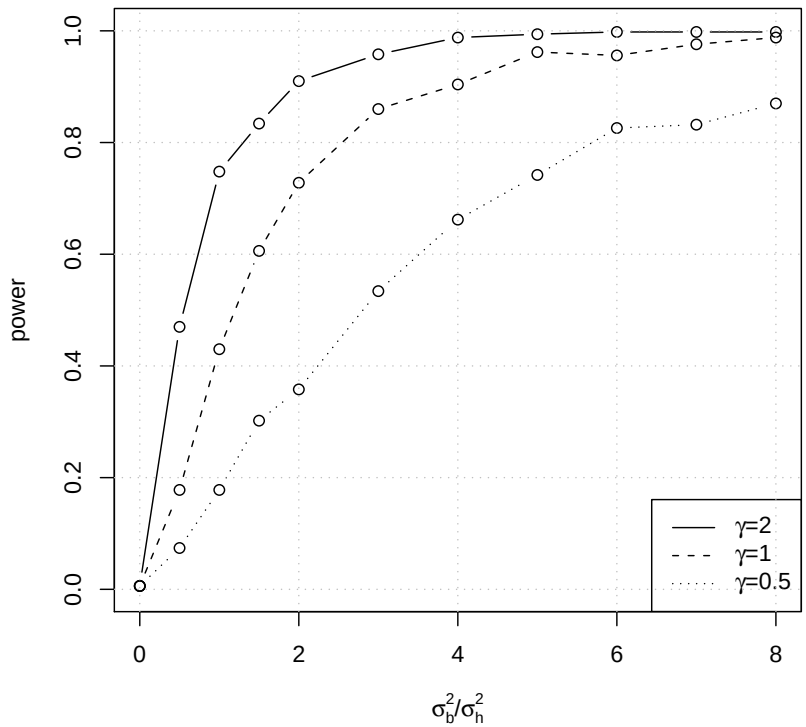
Table : p_D and DIC for competing models in the analysis of Ontario lung cancer incidence.

Type		Model	p_D	DIC
Shared	M1	$\log(\mu_{im}) = \alpha_m + \log E_{im} + b_i + h_{im}$ $\log(\mu_{if}) = \alpha_f + \log E_{if} + \gamma \cdot b_i + h_{if}$	56.5	725.9
	M2	$\log(\mu_{im}) = \alpha_m + \log E_{im} + b_i$ $\log(\mu_{if}) = \alpha_f + \log E_{if} + \gamma \cdot b_i + h_{if}$	56.5	727.5
	M3	$\log(\mu_{im}) = \alpha_m + \log E_{im} + b_i + h_{im}$ $\log(\mu_{if}) = \alpha_f + \log E_{if} + \gamma \cdot b_i$	56.6	728.3
	M4	$\log(\mu_{im}) = \alpha_m + \log E_{im} + b_i$ $\log(\mu_{if}) = \alpha_f + \log E_{if} + \gamma \cdot b_i$	38.8	764.1
Separate	M5	$\log(\mu_{im}) = \alpha_m + \log E_{im} + b_{im} + h_{im}$ $\log(\mu_{if}) = \alpha_f + \log E_{if} + b_{if} + h_{if}$	68.3	738.2
	M6	$\log(\mu_{im}) = \alpha_m + \log E_{im} + b_{im}$ $\log(\mu_{if}) = \alpha_f + \log E_{if} + b_{if} + h_{if}$	68.8	740.1
	M7	$\log(\mu_{im}) = \alpha_m + \log E_{im} + b_{im} + h_{im}$ $\log(\mu_{if}) = \alpha_f + \log E_{if} + b_{if}$	69.4	741.6
	M8	$\log(\mu_{im}) = \alpha_m + \log E_{im} + b_{im}$ $\log(\mu_{if}) = \alpha_f + \log E_{if} + b_{if}$	70.1	743.9

Power of the Test of Common Spatial Structure

Hypothesis test: $H_0 : \gamma = 0 \quad \text{versus} \quad H_1 : \gamma \neq 0\,.$





Precision Gains Through Joint Outcome Modeling

Table : The average absolute relative bias (ABIAS) of estimated risks, as well as their average standard deviation (ASE) and average root mean squared error (ARMSE), along with average exceedance probability (APREX) for regions with true relative risks greater than one. The expected disease counts are scaled by the inverse of δ. The true value of σ_h^2 is 0.01.

		$\sigma_b^2 = 0.1$		$\sigma_b^2 = 0.5$		$\sigma_b^2 = 1$	
		Joint	Separate	Joint	Separate	Joint	Separate
$\delta = 1$	ABIAS	0.073	0.073	0.135	0.135	0.215	0.214
	ASE	0.044	0.047	0.044	0.048	0.043	0.047
	ARMSE	0.099	0.099	0.174	0.175	0.301	0.303
	APREX	0.767	0.757	0.790	0.783	0.779	0.776
$\delta = 50$	ABIAS	0.113	0.122	0.155	0.185	0.190	0.222
	ASE	0.132	0.136	0.190	0.215	0.212	0.252
	ARMSE	0.196	0.205	0.296	0.331	0.399	0.442
	APREX	0.567	0.560	0.634	0.614	0.668	0.644
$\delta = 100$	ABIAS	0.130	0.136	0.199	0.230	0.238	0.288
	ASE	0.143	0.145	0.221	0.242	0.260	0.297
	ARMSE	0.219	0.225	0.349	0.385	0.462	0.513
	APREX	0.528	0.523	0.590	0.575	0.615	0.592

Modeling for Zero-Inflated Data

Zero Inflation

- The presence of excess zeros is a special case of overdispersion,
- Excess zeros may arise from unsuitable habitat or immunity of individuals,
- Difference in excess zeros and sampling zeros.

Zero-inflated Models for Count Data

- Mixture model and two-part model.

Zero-inflated Models for Correlated Data

- GEE or random effect model.

Zero-inflated Models for Spatially Correlated Data

- Spatial correlation in the counts and in the probabilities for subjects coming from the component of excess zeros.

Modeling for Zero-Inflated Data

Zero Inflation

- The presence of excess zeros is a special case of overdispersion,
- Excess zeros may arise from unsuitable habitat or immunity of individuals,
- Difference in excess zeros and sampling zeros.

Zero-inflated Models for Count Data

- Mixture model and two-part model.

Zero-inflated Models for Correlated Data

- GEE or random effect model.

Zero-inflated Models for Spatially Correlated Data

- Spatial correlation in the counts and in the probabilities for subjects coming from the component of excess zeros.

Modeling for Zero-Inflated Data

Zero Inflation

- The presence of excess zeros is a special case of overdispersion,
- Excess zeros may arise from unsuitable habitat or immunity of individuals,
- Difference in excess zeros and sampling zeros.

Zero-inflated Models for Count Data

- Mixture model and two-part model.

Zero-inflated Models for Correlated Data

- GEE or random effect model.

Zero-inflated Models for Spatially Correlated Data

- Spatial correlation in the counts and in the probabilities for subjects coming from the component of excess zeros.

Modeling for Zero-Inflated Data

Zero Inflation

- The presence of excess zeros is a special case of overdispersion,
- Excess zeros may arise from unsuitable habitat or immunity of individuals,
- Difference in excess zeros and sampling zeros.

Zero-inflated Models for Count Data

- Mixture model and two-part model.

Zero-inflated Models for Correlated Data

- GEE or random effect model.

Zero-inflated Models for Spatially Correlated Data

- Spatial correlation in the counts and in the probabilities for subjects coming from the component of excess zeros.

Suppose the response variable Y_{ij} is distributed as

$$Y_{ij}|z_{ij} = \begin{cases} 0 & \text{if } z_{ij} = 1, \\ \text{Poisson}(\mu_{ij}) & \text{if } z_{ij} = 0 \end{cases}$$

The probability distribution functions are

$$Pr(Y_{ij} = y_{ij}) = \begin{cases} \pi_{ij} + (1 - \pi_{ij})e^{-\mu_{ij}} & \text{if } y_{ij} = 0, \\ (1 - \pi_{ij})\frac{e^{-\mu_{ij}}\mu_{ij}^{y_{ij}}}{y_{ij}!} & \text{if } y_{ij} > 0 \end{cases}.$$

The common spatial factor model is

$$\log(\mu_{ij}) = \alpha_j + \gamma_j b_i + h_{ij}, \quad \text{logit}(\pi_{ij}) = \zeta_j + \omega_j d_i \,,$$

where

$$\begin{aligned}
\mathbf{b} &= (b_1, \cdots, b_n)^T \sim \text{MVN}(\mathbf{0}, \sigma_b^2(\mathbf{D} - \mathbf{W})^{-1}), \\
\mathbf{d} &= (d_1, \cdots, d_n)^T \sim \text{MVN}(\mathbf{0}, \sigma_d^2(\mathbf{D} - \mathbf{W})^{-1}), \\
\mathbf{h}_j &= (h_{1j}, \cdots, h_{nj})^T \sim \text{MVN}(\mathbf{0}, \sigma_{h_j}^2 \mathbf{I})
\end{aligned}$$

The joint posterior distribution of the parameters is:

$$p(\boldsymbol{\alpha}, \boldsymbol{\zeta}, \mathbf{b}, \mathbf{d}, \mathbf{h}, \boldsymbol{\gamma}, \boldsymbol{\omega}, \sigma_b^2, \sigma_d^2 | \mathbf{Y}) \propto \quad Ł(\mathbf{Y}|\boldsymbol{\alpha}, \boldsymbol{\zeta}, \mathbf{b}, \mathbf{d}, \mathbf{h}, \boldsymbol{\gamma}, \boldsymbol{\omega}) p(\boldsymbol{\alpha}) p(\boldsymbol{\zeta}) p(\boldsymbol{\gamma}) p(\boldsymbol{\omega})$$
$$p(\mathbf{b}|\sigma_b^2) p(\mathbf{d}|\sigma_d^2) p(\mathbf{h}|\sigma_h^2) p(\sigma_b^2) p(\sigma_d^2) p(\sigma_h^2) \,, \quad (1)$$

The first term is the conditional likelihood,

$$L(\mathbf{Y}|\boldsymbol{\alpha}, \boldsymbol{\zeta}, \mathbf{b}, \mathbf{d}, \mathbf{h}, \boldsymbol{\gamma}, \boldsymbol{\omega}) \propto \prod_{i=1}^{n} \prod_{j=1}^{J} \left[I(y_{ij} = 1) \left\{ \pi_{ij} + (1 - \pi_{ij}) e^{\mu_{ij}} \right\} \right.$$
$$\left. + I(y_{ij} = 0) \left\{ (1 - \pi_{ij}) \frac{e^{-\mu_{ij}} \mu_{ij}^{y_{ij}}}{y_{ij}!} \right\} \right] , \quad (2)$$

To avoid computational instability, normal priors can be assigned on $\boldsymbol{\alpha}$, $\boldsymbol{\zeta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\omega}$ with a moderately large variance and uniform distribution with again moderately large variance for σ_b, σ_d and σ_h.

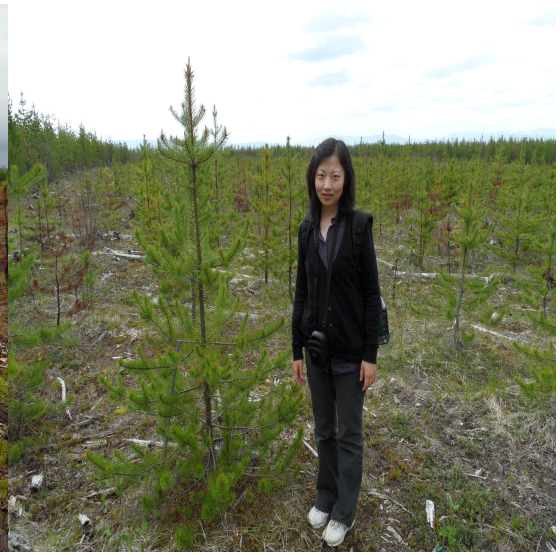
Comandra Blister Rust Study

Comandra blister rust (CBR) is a disease of hard pines that is caused by a fungus growing in the inner bark; CBR infects pines but needs an alternate host plant (AHP) to spread from one pine to another.

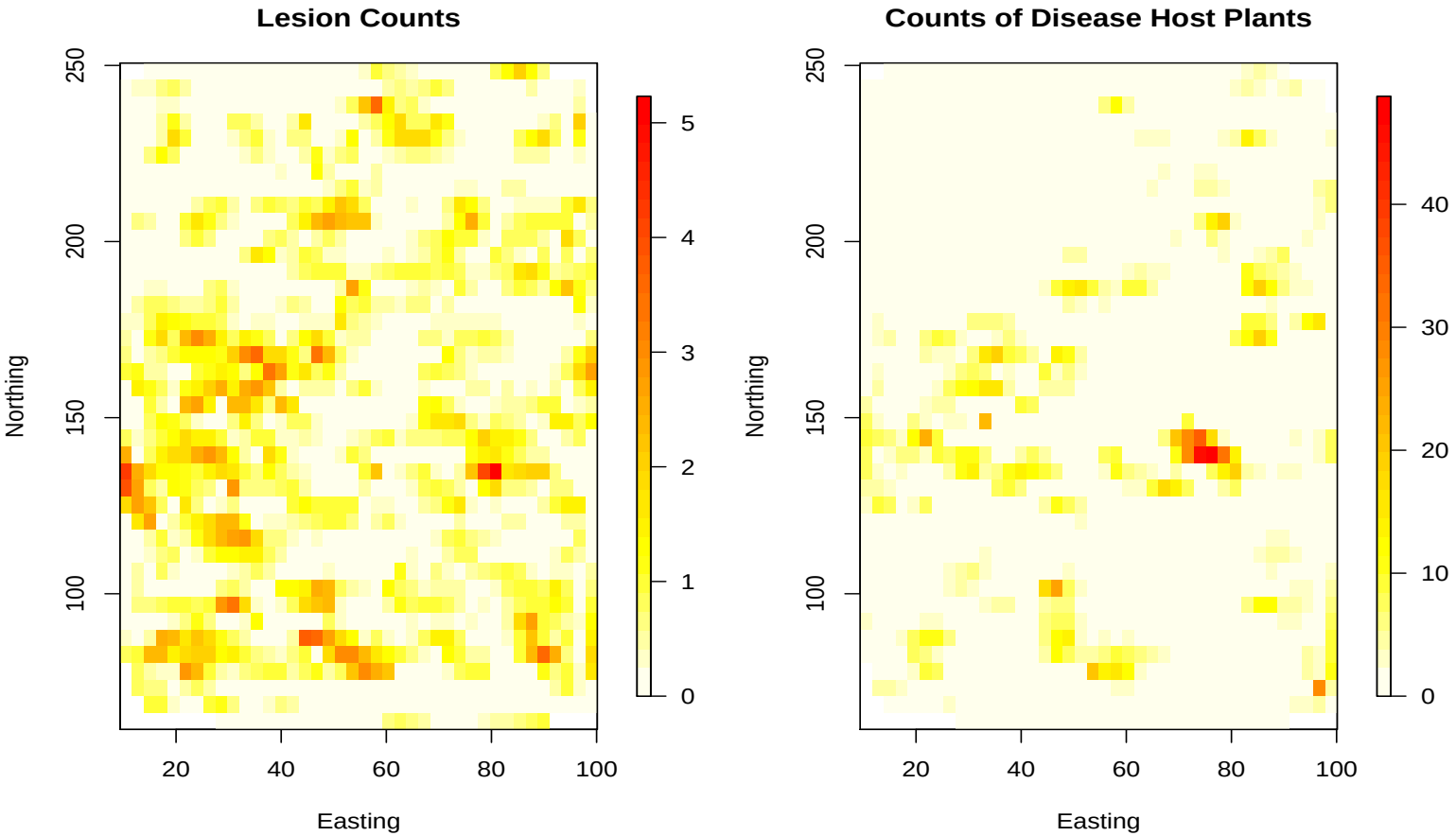
- Plantation of lodgepole pine trees over a 124×64 grid;
- Each grid is 1.5 meters $\times$ 1.5 meters;
- 1000 trees susceptible to CBR infection were randomly sampled over the field;
- In each grid cell, two outcomes:
  (1) counts of lesions on each tree,
  (2) counts of disease host plants in each grid cell.

Comandra Blister Rust Study

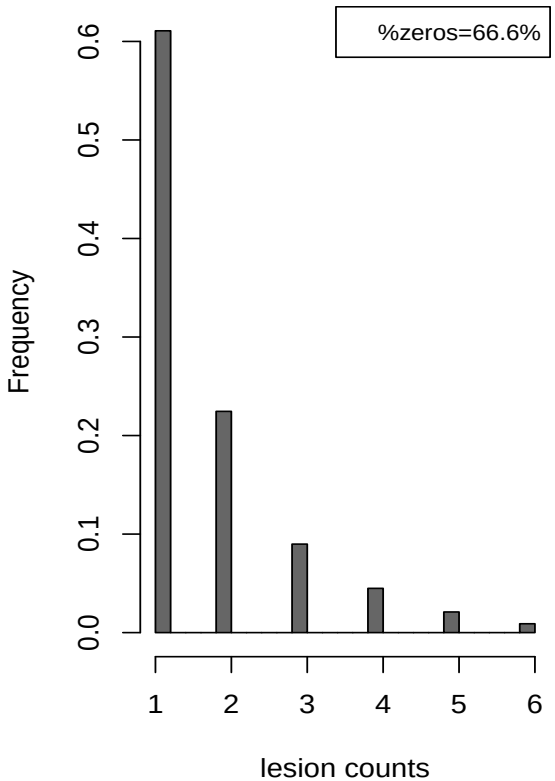


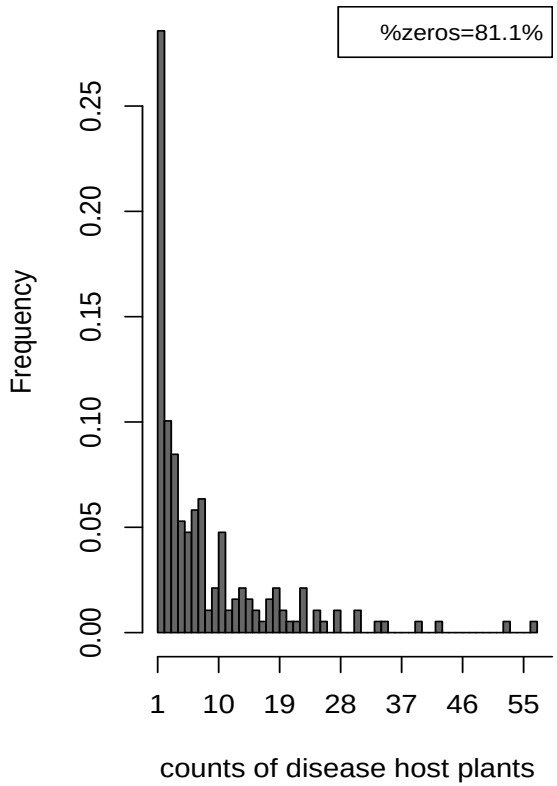



Comandra Blister Rust Study



Comandra Blister Rust Study





Joint Modeling of Tree Infection and Host Plants

Let L and H index counts of lesion and counts of disease host plants, respectively. The zero-inflated common spatial factor model is then,

$$\begin{cases} \log(\mu_{Hi}) = \alpha_H + b_i + h_{Hi} & \text{logit}(\pi_{Hi}) = \zeta_H + d_i \\ \log(\mu_{Li}) = \alpha_L + \gamma \cdot b_i + h_{Li} & \text{logit}(\pi_{Li}) = \zeta_L + \omega \cdot d_i \end{cases}$$

where

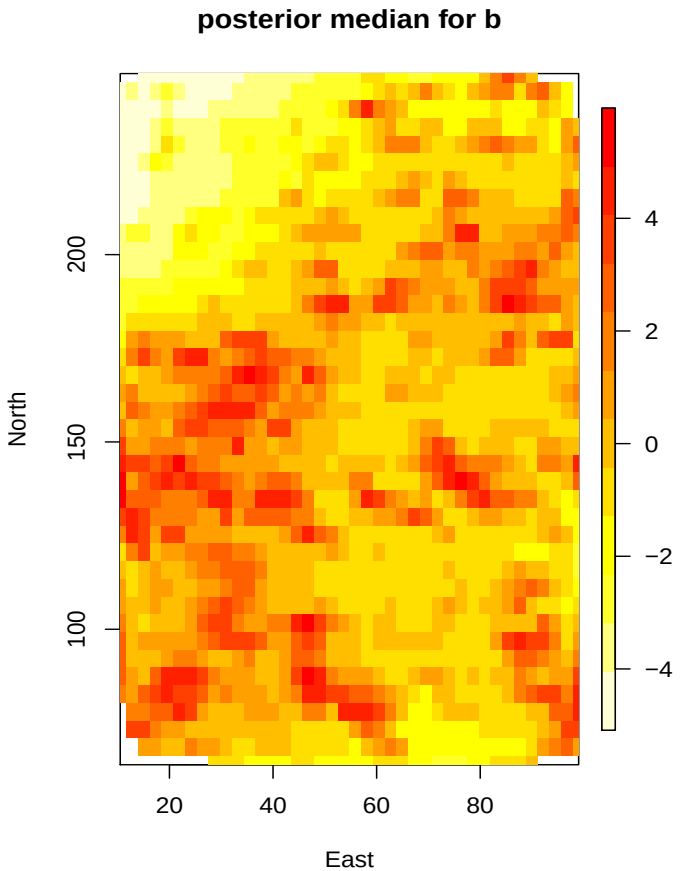
$$\begin{aligned} \mathbf{b} &= (b_1, \cdots, b_n)^T \sim \text{MVN}(\mathbf{0}, \sigma_b^2(\mathbf{D} - \mathbf{W})^{-1}), \\ \mathbf{d} &= (d_1, \cdots, d_n)^T \sim \text{MVN}(\mathbf{0}, \sigma_d^2(\mathbf{D} - \mathbf{W})^{-1}), \\ \mathbf{h}_L &= (h_{1L}, \cdots, h_{nL})^T \sim \text{MVN}(\mathbf{0}, \sigma_{h_L}^2 \mathbf{I}), \\ \mathbf{h}_H &= (h_{1H}, \cdots, h_{nH})^T \sim \text{MVN}(\mathbf{0}, \sigma_{h_H}^2 \mathbf{I}), \end{aligned}$$

Table : Posterior summaries

Parameter	Mean	95% CI
γ	0.25	(0.23, 0.33)
ω	1.42	(0.03, 5.88)
σ_b^2	532.49	(354.53, 793.72)
σ_{hH}^2	0.89	(0.35, 2.57)
σ_{hL}^2	0.20	(0.11, 0.37)

Remark: Spatial similarity of the random process exists across the spatial maps through a latent random effect and zero mass components of the two distributions are also correlated though a latent spatially varying process.

Posterior Medians of the Shared Random Effects



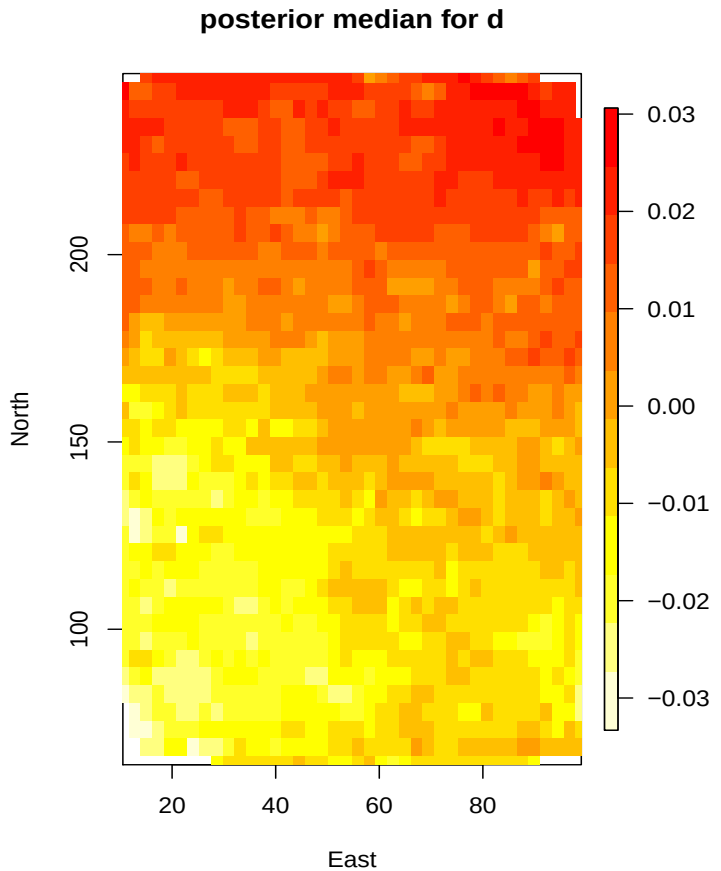


Table : pD and DIC for competing models

Type	Model	Poisson distribution	Excess Zero	pD	DIC
Shared	M1	$\log(\mu_{iH}) = \alpha_H + b_i + h_{iH}$ $\log(\mu_{iL}) = \alpha_L + \gamma_b \cdot b_i + h_{iL}$	$\text{logit}(\pi_{iH}) = \zeta_H + d_i$ $\text{logit}(\pi_{iL}) = \zeta_L + \gamma_d \cdot d_i$	530.0	3457.5
	M2	$\log(\mu_{iH}) = \alpha_H + b_i + h_{iH}$ $\log(\mu_{iL}) = \alpha_L + \gamma_b \cdot b_i + h_{iL}$	$\text{logit}(\pi_{iH}) = \zeta_H$ $\text{logit}(\pi_{iL}) = \zeta_L$	534.8	3462.6
	M3	$\log(\mu_{iH}) = \alpha_H + b_i$ $\log(\mu_{iL}) = \alpha_L + \gamma_b \cdot b_i$	$\text{logit}(\pi_{iH}) = \zeta_H + d_i$ $\text{logit}(\pi_{iL}) = \zeta_L + \gamma_d \cdot d_i$	461.0	3505.5
	M4	$\log(\mu_{iH}) = \alpha_H + b_i$ $\log(\mu_{iL}) = \alpha_L + \gamma_b \cdot b_i$	$\text{logit}(\pi_{iH}) = \zeta_H$ $\text{logit}(\pi_{iL}) = \zeta_L$	462.4	3508.7
Separate	M5	$\log(\mu_{iH}) = \alpha_H + b_{iH} + h_{iH}$ $\log(\mu_{iL}) = \alpha_L + b_{iL} + h_{iL}$	$\text{logit}(\pi_{iH}) = \zeta_H + d_{iH}$ $\text{logit}(\pi_{iL}) = \zeta_L + d_{iL}$	658.0	3645.0
	M6	$\log(\mu_{iH}) = \alpha_H + b_{iH} + h_{iH}$ $\log(\mu_{iL}) = \alpha_L + b_{iL} + h_{iL}$	$\text{logit}(\pi_{iH}) = \zeta_H$ $\text{logit}(\pi_{iL}) = \zeta_L$	654.9	3659.8
	M7	$\log(\mu_{iH}) = \alpha_H + b_{iH}$ $\log(\mu_{iL}) = \alpha_L + b_{iL}$	$\text{logit}(\pi_{iH}) = \zeta_H + d_{iH}$ $\text{logit}(\pi_{iL}) = \zeta_L + d_{iL}$	513.6	3721.7
	M8	$\log(\mu_{iH}) = \alpha_H + b_{iH}$ $\log(\mu_{iL}) = \alpha_L + b_{iL}$	$\text{logit}(\pi_{iH}) = \zeta_H$ $\text{logit}(\pi_{iL}) = \zeta_L$	547.2	3723.7

Remarks

- The common spatial factor model can be used to identify the common spatial structures across outcomes and it also offers some improvement in efficiency of estimating relative risks.
- Correlations of the two components of the zero-inflated model would be useful.
- Spatio-temporal extension of the common spatial factor model would be interesting.
- Gains may be obtained by assuming the shared frailty term is spatially uncorrelated, when it is not clear what neighborhood spatial structure is appropriate.
- Shared frailty models have found utility in many applications.

Remarks

- The common spatial factor model can be used to identify the common spatial structures across outcomes and it also offers some improvement in efficiency of estimating relative risks.
- Correlations of the two components of the zero-inflated model would be useful.
- Spatio-temporal extension of the common spatial factor model would be interesting.
- Gains may be obtained by assuming the shared frailty term is spatially uncorrelated, when it is not clear what neighborhood spatial structure is appropriate.
- Shared frailty models have found utility in many applications.

Remarks

- The common spatial factor model can be used to identify the common spatial structures across outcomes and it also offers some improvement in efficiency of estimating relative risks.
- Correlations of the two components of the zero-inflated model would be useful.
- Spatio-temporal extension of the common spatial factor model would be interesting.
- Gains may be obtained by assuming the shared frailty term is spatially uncorrelated, when it is not clear what neighborhood spatial structure is appropriate.
- Shared frailty models have found utility in many applications.

Remarks

- The common spatial factor model can be used to identify the common spatial structures across outcomes and it also offers some improvement in efficiency of estimating relative risks.
- Correlations of the two components of the zero-inflated model would be useful.
- Spatio-temporal extension of the common spatial factor model would be interesting.
- Gains may be obtained by assuming the shared frailty term is spatially uncorrelated, when it is not clear what neighborhood spatial structure is appropriate.
- Shared frailty models have found utility in many applications.

Remarks

- The common spatial factor model can be used to identify the common spatial structures across outcomes and it also offers some improvement in efficiency of estimating relative risks.
- Correlations of the two components of the zero-inflated model would be useful.
- Spatio-temporal extension of the common spatial factor model would be interesting.
- Gains may be obtained by assuming the shared frailty term is spatially uncorrelated, when it is not clear what neighborhood spatial structure is appropriate.
- Shared frailty models have found utility in many applications.


NSERC
CRSNG


Western
UNIVERSITY · CANADA


GEOIDE


SFU
SIMON FRASER UNIVERSITY
THINKING OF THE WORLD


Ministry of Forests
and Range