Mixture Models for Spatio-Temporal Multi-State Processes

Farouk Nathoo & Charmaine Dean

`nathoo@stat.sfu.ca` and `dean@stat.sfu.ca`

Department of Statistics and Actuarial Science

Simon Fraser University

Introduction

- Many scientific studies involve the observation of a group of subjects over time
- When repeated measurements are taken on each subject the resulting data are referred to as longitudinal data

Introduction

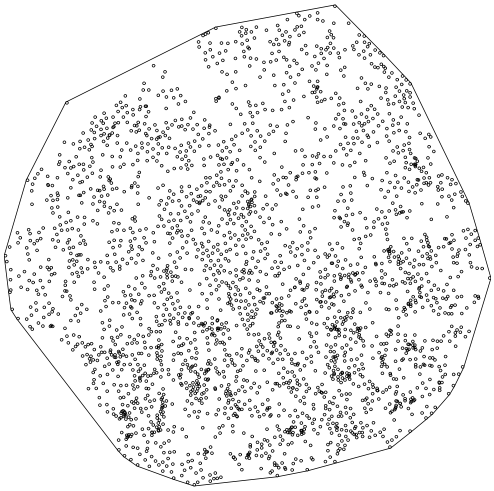
- Well established methods, accounting for temporal correlation and trends, already exist for the analysis of longitudinal data. These include:
  - generalized linear mixed models
  - hidden Markov models
  - counting process models
  - multi-state models

- Existing methods, developed for the health context, typically assume that subjects under observations are independent

- In forestry, statistical methodology for the analysis of longitudinal data should allow for spatial correlation between subjects

- Example: In a study of recurrent weevil infestation, pine trees within a plantation are examined annually for the presence or absence of infection → longitudinal binary data

White Pine Tree Plantation



Introduction

- Use both continuous and discrete mixture models
- Incorporate spatially correlated random effects into existing models for longitudinal data
- Focus in this talk: spatial mixed multi-state models

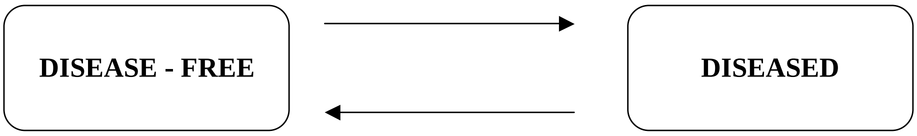
Outline

1. Introduction to multi-state models
2. Discrete-time spatial mover-stayer model for recurrent weevil infestation
3. Continuous-time spatial multi-state models
4. Extensions

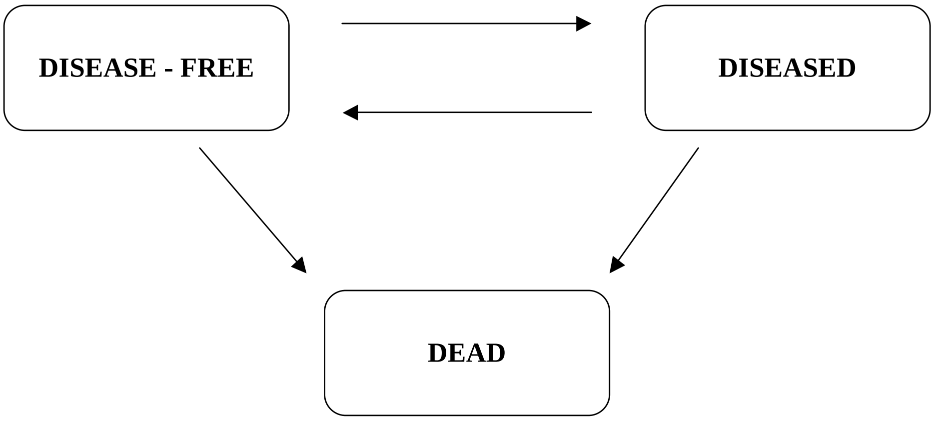
Multi-State Models: A Brief Introduction

- Commonly used in the analysis of longitudinal data.
- At any point in time subjects occupy one of a discrete set of states.
- States can represent the health status of an individual or summarize states of disease activity.
- Change of state is called a transition. For example subjects can make transitions between diseased and disease-free states.

Two-State Model for Disease



Three-State Illness – Death Model



Multi-State Models

- We are interested in modelling the process governing transitions.

- If we observe a discrete time, k-state process, $Y(t), t = 1, 2, ...,$ we are often interested in transition probabilities, for example the 1-step probabilities:

  $$p_{ij}(t) = Pr(Y(t) = j|Y(t-1) = i), \; i, j \in \{1, ..., k\}$$

- For a continuous time process $Y(t), t \in [0, \tau]$, we are interested in analogous quantities known as the transition intensity functions:

  $$\lambda_{ij}(t) = \lim_{h \downarrow 0} \frac{Pr(Y(t) = j|Y(t-h) = i, H_{t-h})}{h}$$

  $$i, j \in \{1, ..., k\}$$

Multi-State Models

- Common simplifying assumption is the Markov assumption where one assumes the entire history of the process up to time t is captured by the current state occupied at time t.

- While many complex processes may not exhibit Markov dependence, Markov processes can be used as building blocks in a hierarchical setting to specify more complex models.

Multi-State Models

- When a large number of subjects are observed, there can be a great deal of heterogeneity in the transition processes corresponding to different subjects. This can be accommodated in several ways:

  1. Regression models, Muenz and Rubenstein (1985)

  2. Independent random effect models, Cook and Ng (1997), Albert and Waclawiw (1998)

  3. Finite mixture models, Frydman (1984), Fuchs and Greenhouse (1988), Cook et al. (2002)

- These models are more sophisticated but still assume independence between subjects.

Weevil Infestation in B.C. Forests

- Study involves a plantation of pine trees roughly 22,000 m^2 in size.
- Roughly 3000 pine trees susceptible to weevil infection.
- Each tree was examined at regular annual intervals for the presence/absence of weevil infection.
- Trees are therefore observed in one of two states, either infected or uninfected and make transitions between the two states.
- Objective: describe the transition process between states and characterize variation in this process over space and time.

Weevil Infestation in B.C. Forests

- Weevil attacks may occur in clusters throughout the region → spatial correlation in the year-to-year disease status of trees.

- In addition 27% of trees were never observed in the infected state throughout the course of the study.

- It was thought that some trees may be ”resistant” to infection. These are trees which, for unknown reasons, have extremely low probability of infection and are thus always observed in the uninfected state.

A Model for Tree Infection

- Let state 1 denote the infected state and state 0 denote the uninfected state.

- The state occupied by the i^{th} tree at year t is denoted by a binary variate $y_i(t)$.

- Let $\mathbf{x_i(t)}$ denote a corresponding vector of covariates.

- The response obtained from the i^{th} tree is therefore a vector of binary values

  $$\mathbf{y_i} = (y_i(0), ..., y_i(6))'$$

  indicating the sequence of states occupied.

A Model for Tree Infection

- We specify a hierarchical model where, at the first level of the model, we assume each response vector, $\mathbf{y_i}$, is independently drawn from a two component mixture model:

$$\mathbf{y_i} \sim \begin{cases} f_{1_i}(\mathbf{Y_i}) & \text{with } p_M \\ f_{0_i}(\mathbf{Y_i}) & \text{with } 1 - p_M \end{cases}$$

- $f_{0_i}(\mathbf{Y_i})$ is the density of a degenerate distribution placing all its mass on the zero vector

$$f_{0_i}(\mathbf{Y_i}) = I\{\mathbf{Y_i} = \mathbf{0}\}$$

A Model for Tree Infection

- $f_{1_i}(\mathbf{Y_i})$ is the density of a 1^{st} order, 2-state Markov chain with initial probability:

  $$p_{I_i} = Pr\{y_i(0) = 1\}$$

  and transition probabilities:

  $$p_{01_i}(t) = Pr\{y_i(t) = 1 | y_i(t-1) = 0\}$$

  $$p_{10_i}(t) = Pr\{y_i(t) = 0 | y_i(t-1) = 1\}$$

  $$t = 1, ..., 6.$$

A Model for Tree Infection

- We model transition probabilities with logistic regression specifications:

$$logit\{p_{01_i}(t)\} = \boldsymbol{\beta_0}'\mathbf{x_i(t)} + g_0(t, \boldsymbol{\alpha_0}) + b_{0_i}$$

$$logit\{p_{10_i}(t)\} = \boldsymbol{\beta_1}'\mathbf{x_i(t)} + g_1(t, \boldsymbol{\alpha_1}) + b_{1_i}$$

A Model for Tree Infection

- We model transition probabilities with logistic regression specifications:

$$logit\{p_{01_i}(t)\} = \boldsymbol{\beta_0}'\mathbf{x_i(t)} + g_0(t, \boldsymbol{\alpha_0}) + b_{0_i}$$

$$logit\{p_{10_i}(t)\} = \boldsymbol{\beta_1}'\mathbf{x_i(t)} + g_1(t, \boldsymbol{\alpha_1}) + b_{1_i}$$

- $\mathbf{x_i(t)} = (1, D_i, A_i(t-1))'$

  1. D_i a local measure of tree density
  2. $A_i(t-1)$ a local measure of attack density at time $t-1$

A Model for Tree Infection

- We model transition probabilities with logistic regression specifications:

$$logit\{p_{01_i}(t)\} = \boldsymbol{\beta_0}'\mathbf{x_i(t)} + g_0(t, \boldsymbol{\alpha_0}) + b_{0_i}$$

$$logit\{p_{10_i}(t)\} = \boldsymbol{\beta_1}'\mathbf{x_i(t)} + g_1(t, \boldsymbol{\alpha_1}) + b_{1_i}$$

- $\mathbf{x_i(t)} = (1, D_i, A_i(t-1))'$

  1. D_i a local measure of tree density

  2. $A_i(t-1)$ a local measure of attack density at time $t-1$

A Model for Tree Infection

- We model transition probabilities with logistic regression specifications:

$$logit\{p_{01_i}(t)\} = \boldsymbol{\beta_0}'\mathbf{x_i(t)} + g_0(t, \boldsymbol{\alpha_0}) + b_{0_i}$$

$$logit\{p_{10_i}(t)\} = \boldsymbol{\beta_1}'\mathbf{x_i(t)} + g_1(t, \boldsymbol{\alpha_1}) + b_{1_i}$$

- $\mathbf{x_i(t)} = (1, D_i, A_i(t-1))'$

1. D_i a local measure of tree density
2. $A_i(t-1)$ a local measure of attack density at time $t-1$

A Model for Tree Infection

- We model transition probabilities with logistic regression specifications:

$$logit\{p_{01_i}(t)\} = \boldsymbol{\beta_0}'\mathbf{x_i(t)} + g_0(t, \boldsymbol{\alpha_0}) + b_{0_i}$$

$$logit\{p_{10_i}(t)\} = \boldsymbol{\beta_1}'\mathbf{x_i(t)} + g_1(t, \boldsymbol{\alpha_1}) + b_{1_i}$$

- $g_0(t, \boldsymbol{\alpha_0})$ and $g_1(t, \boldsymbol{\alpha_1})$ describe overall temporal trends in transitions - use cubic B-splines

A Model for Tree Infection

- We model transition probabilities with logistic regression specifications:

$$logit\{p_{01_i}(t)\} = \boldsymbol{\beta_0}'\mathbf{x_i}(\mathbf{t}) + g_0(t, \boldsymbol{\alpha_0}) + b_{0_i}$$

$$logit\{p_{10_i}(t)\} = \boldsymbol{\beta_1}'\mathbf{x_i}(\mathbf{t}) + g_1(t, \boldsymbol{\alpha_1}) + b_{1_i}$$

- b_{0_i} and b_{1_i} are tree specific random effects

A Model for Tree Infection

- Two vectors of random $\boldsymbol{b_0} = (b_{0_1}, ..., b_{0_N})'$ and $\boldsymbol{b_1} = (b_{1_1}, ..., b_{1_N})'$.

- We assume each vector is independently drawn from a conditionally autoregressive $CAR(\sigma)$ model: $\mathbf{b_l} \overset{ind}{\sim} CAR(\sigma_l), \; l = 0, 1$

- The CAR model is a Markov random field model where the joint distribution for a collection of spatially correlated random variables is determined through a set of local specifications

- Have found wide application in disease mapping and image analysis

A Model for Tree Infection

- To model the random effects we employ a Gaussian $CAR(\sigma)$ model where the joint distribution for a random vector $\mathbf{b} = (b_1, ..., b_n)$ is specified through the conditional distributions:

$$b_i | \mathbf{b_{j \neq i}} \sim Normal(\mu_i, \sigma_i^2)$$

$$\mu_i = \frac{\sum_{j \neq i} w_{ij} b_j}{\sum_{j \neq i} w_{ij}}, \quad \sigma_i^2 = \frac{\sigma^2}{\sum_{j \neq i} w_{ij}}$$

- Examples of w_{ij}:

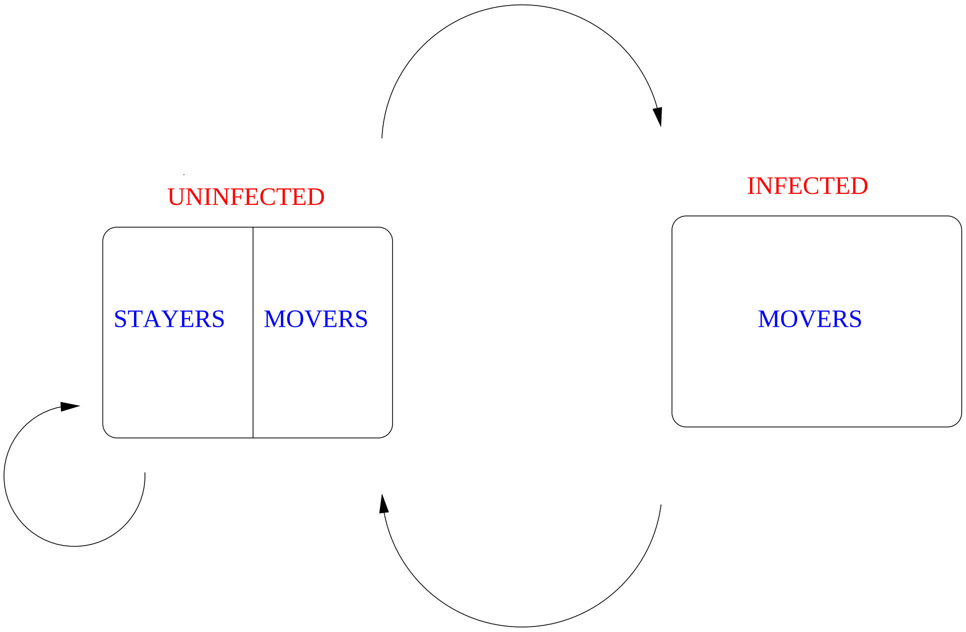
1. $w_{ij} = \begin{cases} 1 & \text{if } d_{ij} \leq d \\ 0 & \text{if } d_{ij} > d \end{cases}$

2. $w_{ij} = \exp\left(-d_{ij}\right)$

A Model for Tree Infection

- Our 2-state model allows for a sub-population of resistant trees - stayers.

- Trees which can make transitions do so according to a Markov chain incorporating spatially correlated random effects - movers.



Inference

- We approach from both classical and Bayesian perspectives:

  1. Marginal Likelihood Function

  $$L(\boldsymbol{\Theta}, \mathbf{Y}) = E_{\boldsymbol{b_0},\boldsymbol{b_1}}[\prod_{i=1}^{N} f(\mathbf{y_i}|\boldsymbol{b_1}, \boldsymbol{b_0})]$$

  2. Posterior Distribution

  $$\pi(\boldsymbol{\Theta}, \mathbf{b_0}, \mathbf{b_1}|\mathbf{Y}) \propto \left[\prod_{i=1}^{N} f(\mathbf{y_i}|\boldsymbol{b_1}, \boldsymbol{b_0}) \right]$$

  $$\times f(\boldsymbol{b_0}|\sigma_0) f(\boldsymbol{b_1}|\sigma_1) \pi(\boldsymbol{\Theta})$$

- $\pi(\boldsymbol{\Theta})$ prior distribution for parameters. We use weakly informative priors.

- In both cases we apply model fitting techniques based on Markov Chain Monte Carlo simulation.

Inference

- Bayesian analysis: Draw Markov Chain Monte Carlo samples from posterior distribution.

- Likelihood analysis: Use a Monte Carlo EM algorithm to obtain MLE's.

    1. EM used for missing data problems.

    2. Treat random effects as "missing" data.

    3. The E-step of a standard EM algorithm is not available in closed form for our model.

    4. Approximate the required expectation using Gibbs sampling.

Monte Carlo EM

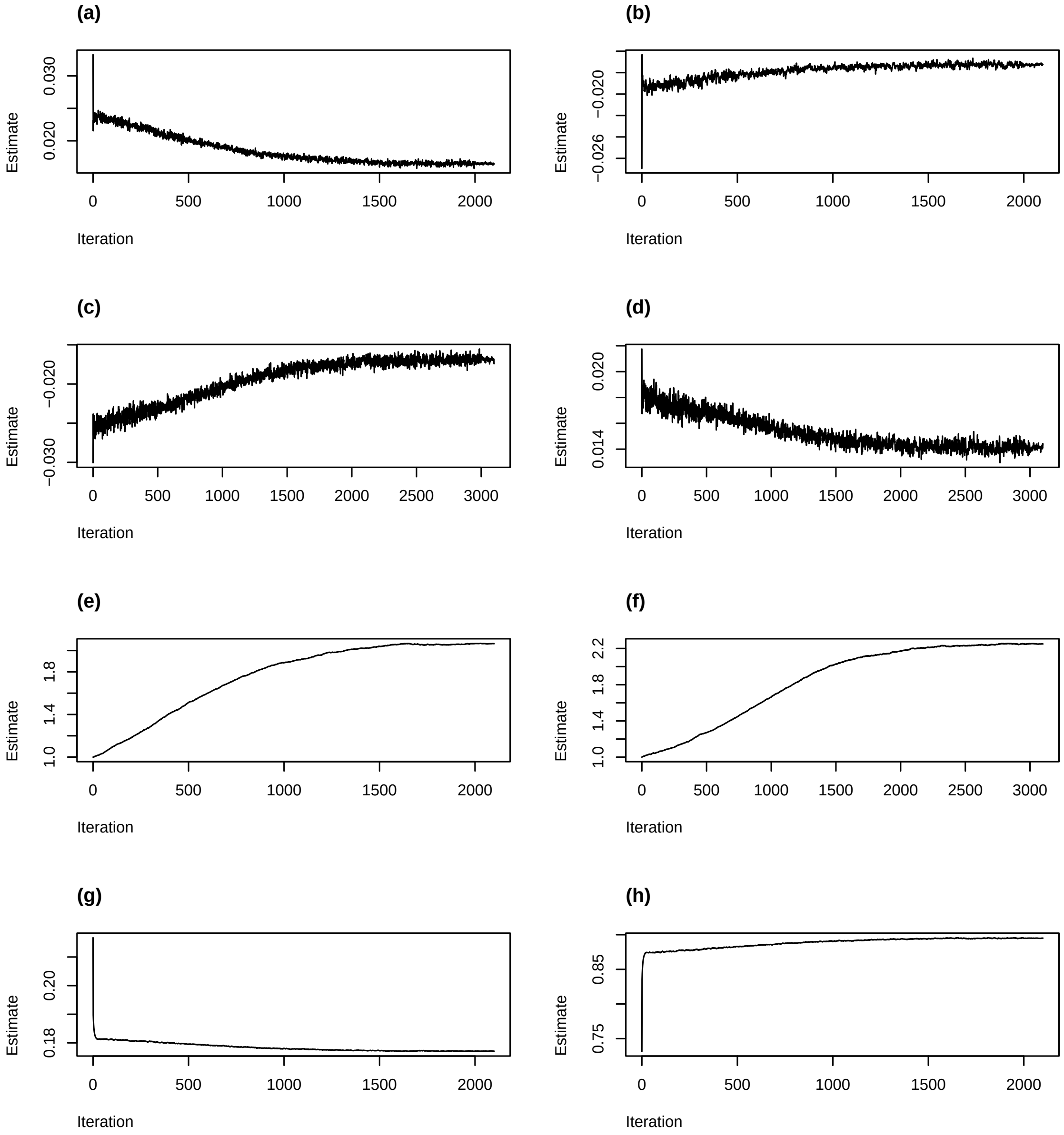
- The MCEM algorithm does not posses the monotonicity properties of the usual EM algorithm.

- Will converge to a neighborhood of the MLE and vary randomly about this neighborhood.

- Convergence can be assessed by examining trace plots of each parameter.

- Standard Errors: use MCMC to approximate the observed information matrix.

MCEM Output for Mover-Stayer Model



MCEM estimates by iteration: (a) β_{0_1} (b) β_{0_2} (c) β_{1_1} (d) β_{1_2} (e) σ_{b_0} (f) σ_{b_1} (g) P_I (h) P_M

Some Results

- Results obtained from Bayesian and Likelihood procedures are very similar.

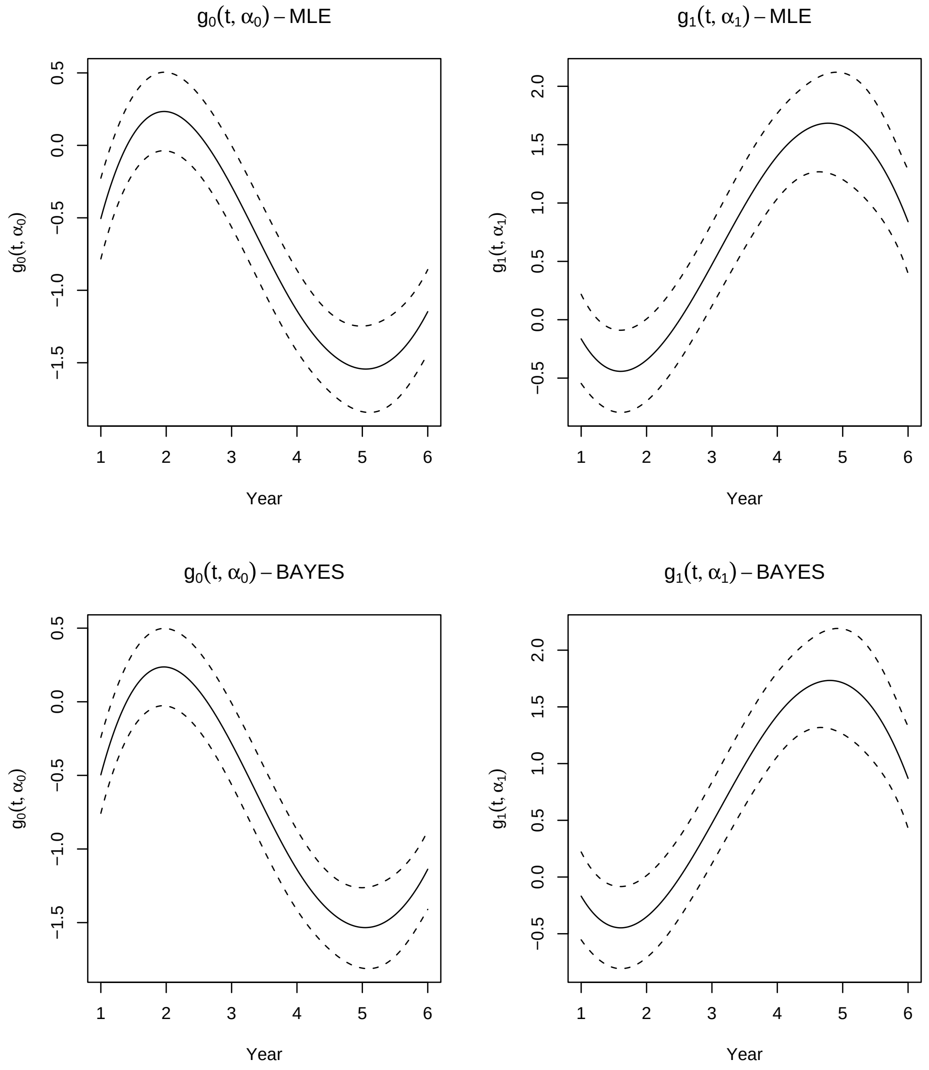
- Mixture allocation probability (proportion of movers):

$$\widehat{p_M} = 0.894\,(0.021) \text{ -- BAYES}$$

$$\widehat{p_M} = 0.895\,(0.018) \text{ -- MLE}$$

- Covariate effect: $A_i(t-1)$ (local measure on infection density)
  - Trees surrounded by a larger number of infected trees at time $t-1$ are more likely to make a transition into the infected state at time t.

Estimated Temporal Trends: $g_0(t, \boldsymbol{\alpha}_0)$, $g_1(t, \boldsymbol{\alpha}_1)$

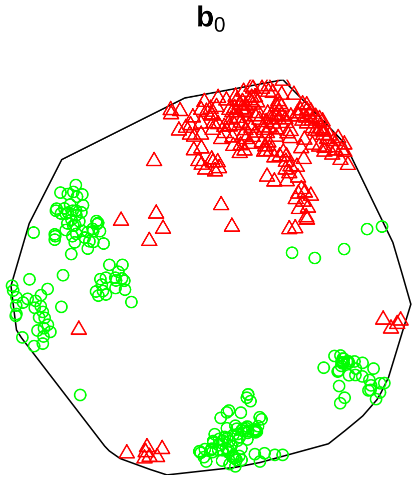


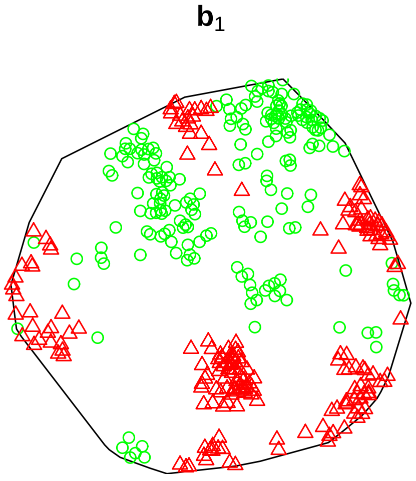
Estimated Random Effects

- We can estimate the random effects - ”spatial residuals”.

- These can be used to obtain maps of the unmodeled heterogeneity → pinpoint areas to search for missing covariates.

Estimated Random Effects: b_0, b_1





Goodness of Fit

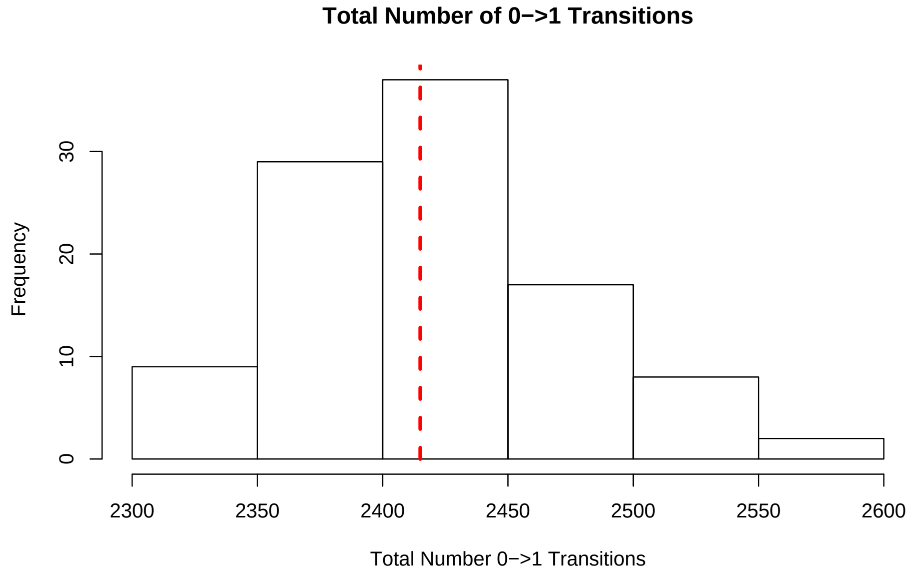
- We use the posterior predictive distribution $pr(\mathbf{Y^{rep}}|\mathbf{Y})$ associated with our model where $\mathbf{Y^{rep}}$ is a hypothetical replicate data set.

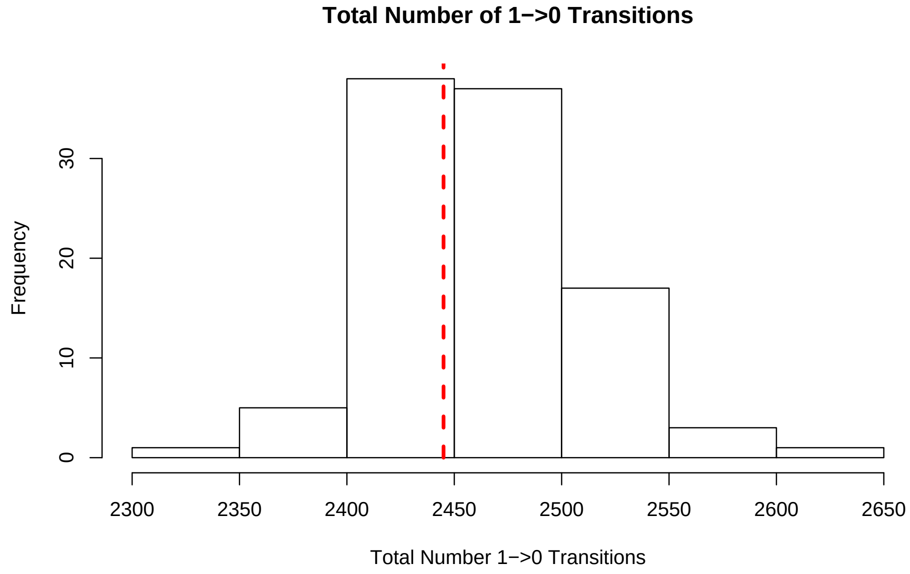
- If the models fits reasonably well, the observed data should look plausible under this distribution.

- Simulate replicate data sets from the posterior predictive distribution $\mathbf{Y_1^{rep}}, \mathbf{Y_2^{rep}}, ..., \mathbf{Y_k^{rep}}$ and compare these to the observed data.

- Any systematic differences between data simulated from our model and the observed data may be an indication of model misfit.

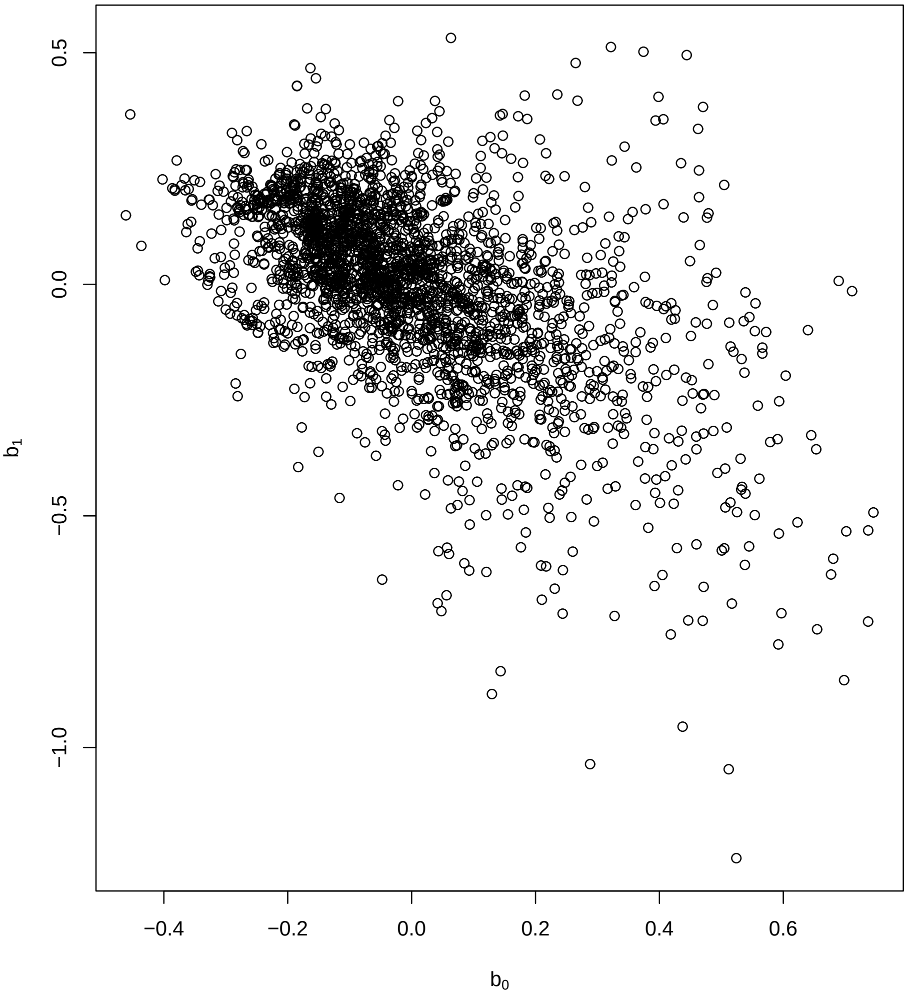
Goodness of Fit





Random Effects - Forest Example

- Negative correlation between b_{0_i} and b_{1_i}



Random Effects - Forest Example

- We seek a joint spatial model for $\boldsymbol{b_0} = ({b_0}_1, ..., {b_0}_N)'$ and $\boldsymbol{b_1} = ({b_1}_1, ..., {b_1}_N)'$
- Employ a bivariate CAR model with conditional specifications for the pair of random effects, $\mathbf{b_i} = (b_{0i}, b_{1i})'$, associated with the i^{th} tree given by:

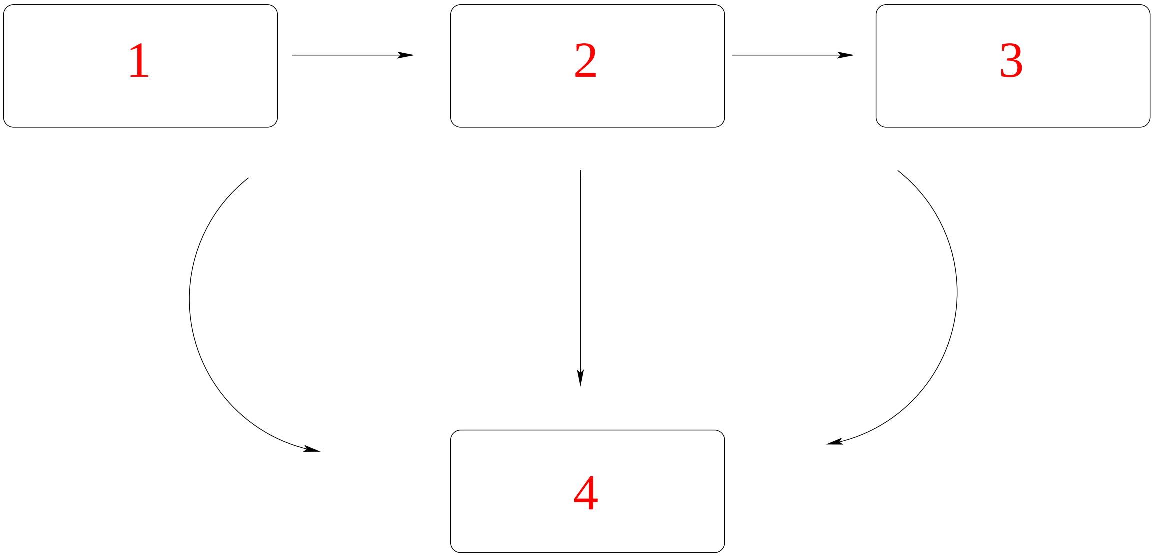
$$\left(\begin{array}{c} b_{0i} \\ b_{1i} \end{array} \right) \Bigg| \left\{ \left(\begin{array}{c} b_{0j} \\ b_{1j} \end{array} \right) \right\}_{j \neq i} \sim BVN(\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i})$$

$$\left(\begin{array}{c} \mu_{0i} \\ \mu_{1i} \end{array} \right) = \sum_j \boldsymbol{W_{ij}} \left(\begin{array}{c} b_{0j} \\ b_{1j} \end{array} \right)$$

- For models with more than two states, multivariate CAR models are defined in a similar manner

Continuous-time Models

- More general continuous-time models for spatially reference event history data



- Describe transition rates between states while accounting for spatial correlation.

- Identify and map spatial variation in transition rates.

Continuous-time Models

- Let $X_l(t)$ represent the state occupied by subject l at time t, $l = 1, ..., n$

- We specify a hierarchical model where, at the first stage of the model, we assume that processes corresponding to different subjects are independent with each following a continuous time Markov Chain

- Let $\lambda_{ijl}(t)$ denote the intensity function associated with transitions from state i to state j for subject l: $\lambda_{ijl}(t) =$

$$\text{lim}_{h\to 0}\frac{Pr\{x_l(t) = j|x_l(t-h) = i\}}{h}$$

Continuous-time Models - Transition Intensity Functions

- We adopt a proportional intensities model given by:

$$\lambda_{ijl}(t) = \lambda_{ij0}(t)\,\mathrm{exp}(\boldsymbol{\beta_{ij}}'\boldsymbol{z_l} + \boldsymbol{b_{ij}}'\boldsymbol{r_l})$$

- $\lambda_{ij0}(t)$: baseline intensity function

- $\boldsymbol{z_l}$: vector of explanatory variables

- $\boldsymbol{\beta_{ij}}$: vector of regression coefficients

- $\boldsymbol{b_{ij}}$ is a vector of random effects allowing for spatial correlation

Continuous-time Models: $\lambda_{ij0}(t)$

- Splines
- Parametric forms, for example Weibull:

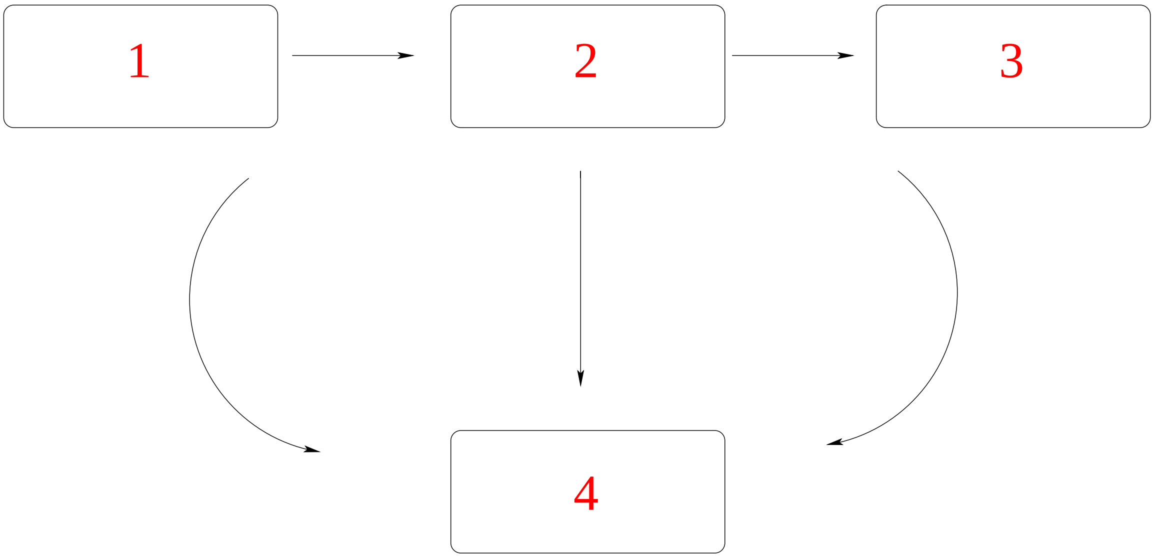
$$\lambda_{ij0}(t) = \rho_{ij} t^{\rho_{ij}-1}$$

with shape parameter $\rho_{ij} > 0$.

Continuous-time Models – Random Effects

- The model contains a separate vector of random effects, $\boldsymbol{b_{ij}}$, for each possible $i \rightarrow j$ transition

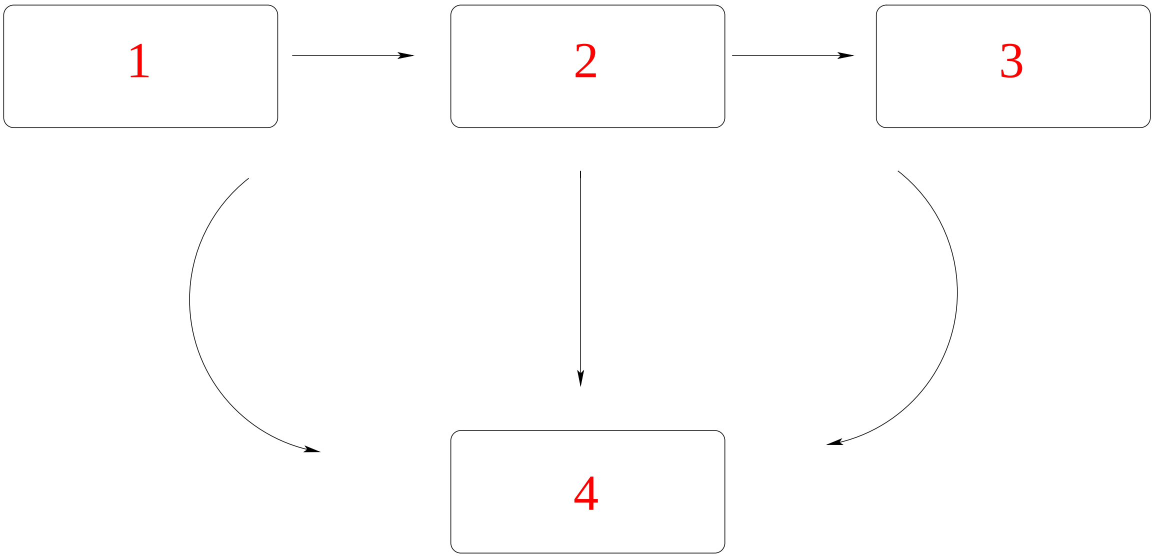
- Model in example would contain five sets of random effects: $\boldsymbol{b_{12}}$, $\boldsymbol{b_{14}}$, $\boldsymbol{b_{23}}$, $\boldsymbol{b_{24}}$ and $\boldsymbol{b_{34}}$



Continuous-time Models – Random Effects

- At the second level of the model we assume that random effects follow a multivariate conditional autoregressive model

- Such a model allows for both spatial correlation in transitions as well as correlation between different transition rates



Continuous-time Models – Model Fitting and Inference

- As before we can employ both Bayesian and frequentist methods based on MCMC sampling

Extensions

- Extensions to the current formulation of the mover-stayer model allowing for spatial clusters of resistant trees

- Extend methods for continuous-time models to accommodate processes under panel observation

Extensions (Links with Jason's work)

- Space-varying coefficient models - do covariate effects vary spatially?

- Random spline models where the coefficients of the spline basis are spatially varying over locations - do peak periods vary spatially?

- Adapative spline models

- Integrate methodology into a GIS/spatial analysis system.