Generalized Linear Models for Spatial Data

Hyoin An and Joshua Radack

April 15, 2021

Outline

1. Introduction
2. Frequentist Method
   - Gotway and Stroup (1997)
3. Generalized Linear Mixed Model (GLMM) Methods
   - Bayesian Approach: Diggle, Tawn, and Moyeed (1998)
   - More recent approach: Bonat and Ribeiro (2015)
4. Application

Outline

Spatial Data

- Spatially-indexed data arise in numerous areas of statistical application
- Methods for spatially dependent data generally assume the data come from a Gaussian Process
- But many kinds of spatially data are not normally distributed (e.g. amount of rainfall, number of species in a region)
- Theory of generalized linear models can be extended to handle cases of non-Gaussian spatially dependent data

Spatial Processes

Let $\{Y(\boldsymbol{x}) : \boldsymbol{x} \in B\}$ be a stochastic process defined on some spatial region B. We characterize the spatial dependence through a covariance function $\mathrm{Cov}(Y(\boldsymbol{x}), Y(\boldsymbol{x}')) = C_Y(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\theta})$ for all $\boldsymbol{x}, \boldsymbol{x}' \in B$.
In practice, we only get to observe $Y(\boldsymbol{x})$ at locations $\boldsymbol{x}_1, ..., \boldsymbol{x}_n \in B$. Let $\boldsymbol{Y} = (Y(\boldsymbol{x}_1), ..., Y(\boldsymbol{x}_n))^T$ be the vector of observations. Define

$$\boldsymbol{\mu} = \mathbb{E}(\boldsymbol{Y}) \quad \Sigma = \mathrm{Cov}(\boldsymbol{Y})$$

to be the mean vector and covariance matrix, respectively. We also assume a link function g relating the mean to a linear combination of covariates,

$$\boldsymbol{\eta} = g(\boldsymbol{\mu}) = D\boldsymbol{\beta},$$

where D is an $n \times p$ design matrix, and variance function $v(\mu_i)$ defined by $\mathrm{Var}(Y_i) = a(\phi_i)v(\mu_i)$.
A key goal of spatial statistics is to make predictions at unobserved spatial locations $\boldsymbol{x}_1^*, ..., \boldsymbol{x}_k^* \in B$.

Outline

Generalized Estimating Equations

Gotway and Stroup (1997) adapted the quasi-likelihood methods for fitting spatially correlated data. The generalized quasi-score is defined as

$$\frac{\partial Q(\boldsymbol{\mu}; \boldsymbol{z})}{\partial \boldsymbol{\mu}} = \Sigma^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}).$$

Differentiating Q with respect to $\boldsymbol{\beta}$ yields the score function

$$\boldsymbol{U} = \Delta^T \Sigma^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}),$$

where Δ is an $n \times p$ matrix with ij element $\frac{\partial \mu_i}{\partial \beta_j}$.

Generalized Estimating Equations

A consistent estimator of $\boldsymbol{\beta}$ can be obtained by setting $\boldsymbol{U} = 0$ using IWLS if we assume the covariance matrix can be written

$$\Sigma = v_{\boldsymbol{\mu}}^{1/2} R(\boldsymbol{\theta}) v_{\boldsymbol{\mu}}^{1/2}$$

where $v_{\boldsymbol{\mu}} = \text{diag}(v(\mu_i))$ are the variance functions and $R(\boldsymbol{\theta})$ is a correlation matrix with parameters $\boldsymbol{\theta}$.
Covariance parameters $\boldsymbol{\theta}$ can be estimated with standard methods, such as least squares with an empirical covariance.

Outline

Bayesian Approach: Spatial GLMM

Suppose $\boldsymbol{Y} = (Y_1, ..., Y_n)$ are non-gaussian observations that are spatially correlated. Then, now we need to incorporate the spatial effects into the GLM.

Recall the random effects models from the lecture:

$$\boldsymbol{\eta}_i = g(\mu_i) = \boldsymbol{d}_i^T \boldsymbol{\beta} + U_i$$

where $g(\cdot)$ is a link function and $\mu_i = \mathbb{E}[Y_i | U_i]$ for $i = 1, ..., n$.

Let the random effect U_i be a spatially-varying process. This model is called a **Spatial Generalized Linear Mixed Model**. (Diggle, Tawn, and Moyeed, 1998)

Bayesian Approach: Spatial GLMM

Diggle, Tawn, and Moyeed (1998) assume the followings:

1. S is a stationary Gaussian process with $\mathbb{E}[S(\boldsymbol{x}_i)] = 0$ and

$$C_{\boldsymbol{\theta}}(S(\boldsymbol{x}_i), S(\boldsymbol{x}_j)) = \text{Cov}[S(\boldsymbol{x}_i), S(\boldsymbol{x}_j)] = \sigma^2 \rho(\boldsymbol{x}_i - \boldsymbol{x}_j),$$

   where $\boldsymbol{\theta}$ is the set of parameters consisting of σ^2 and any parameters related to the correlation structure that exists in S.

2. $Y_1, ..., Y_n$ are conditionally independent given S with distributions

$$f_i(y|S(\boldsymbol{x}_i)) = f(y; \mu_i)$$

   with $\mu_i = \mathbb{E}[Y_i|S(\boldsymbol{x}_i)], i = 1, ..., n$.

3. $g(\mu_i) = \boldsymbol{d}_i^T \boldsymbol{\beta} + S(\boldsymbol{x}_i)$, for some known link function $g(\cdot)$, explanatory variables $\boldsymbol{d}_i = \boldsymbol{d}(\boldsymbol{x}_i)$ and regression parameters $\boldsymbol{\beta}$.

4. $\hat{S}(\boldsymbol{x}) = \mathbb{E}[S(\boldsymbol{x})|\boldsymbol{Y}]$ is the generalized linear predictor for $S(\boldsymbol{x})$.

Bayesian Approach: Conditional densities

- The generalized linear predictor for an unobserved location involves intractable integration, especially when n is large.
- Instead, we can consider MCMC algorithm for approximate evaluation of the quantities needed for prediction.

By assumption 2 on the previous slide, notice that

$$p(\boldsymbol{Y}|\boldsymbol{S}, \beta) = \prod_{i=1}^{n} f(y_i|s_i, \beta),$$

where $f_i(y|s_i, \beta) \equiv f(y; \mu_i)$.

With this, we can write the conditional posterior distributions for each in $(\boldsymbol{\theta}, \boldsymbol{S}, \beta)$, where

$$\boldsymbol{S} = S(\boldsymbol{x}) = (S(\boldsymbol{x}_1), ..., S(\boldsymbol{x}_n))^T = (S_1, ..., S_n)^T$$

Bayesian Approach: Conditional densities

The conditional posterior distributions for $(\boldsymbol{\theta}, \boldsymbol{S}, \beta)$:

$$\pi(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{S}, \beta) = \pi(\boldsymbol{\theta}|\boldsymbol{S}) \propto p(\boldsymbol{S}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

$$\pi(S_j|\boldsymbol{S}_{-j}, \boldsymbol{Y}, \boldsymbol{\theta}, \beta) \propto p(\boldsymbol{Y}|\boldsymbol{S}, \beta)p(S_j|\boldsymbol{S}_{-j}, \boldsymbol{\theta}) = \left\{\prod_{i=1}^{n} f(y_i|s_i, \beta)\right\} p(S_j|\boldsymbol{S}_{-j}, \boldsymbol{\theta})$$

$$\pi(\beta|\boldsymbol{Y}, \boldsymbol{S}, \boldsymbol{\theta}) = \pi(\beta|\boldsymbol{Y}, \boldsymbol{S}) \propto p(\boldsymbol{Y}|\boldsymbol{S}, \beta)p(\beta) = \left\{\prod_{i=1}^{n} f(y_i|s_i, \beta)\right\} p(\beta)$$

Note:

- Since $p(\boldsymbol{S}|\boldsymbol{\theta})$ has a multivariate normal density, $p(S_j|\boldsymbol{S}_{-j}, \boldsymbol{\theta})$ has a univariate Gaussian distribution.
- Diggle et al. (1998) choose $p(\boldsymbol{\theta}) \propto 1, p(\beta) \propto 1$.

Bayesian Approach: MCMC algorithm

Step 1 (Update θ) Generate $\boldsymbol{\theta}' \sim p(\boldsymbol{\theta})$ and accept $\boldsymbol{\theta}'$ with probability

$$\Delta(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min\left\{\frac{p(\boldsymbol{S}|\boldsymbol{\theta}')}{p(\boldsymbol{S}|\boldsymbol{\theta})}, 1\right\}$$

Step 2 (Update $\boldsymbol{S}$) Generate $\boldsymbol{S}' = (S_1', ..., S_n'), S_j' \sim q(S_j, S_j')$, where $q(S_j, S_j') = p(S_j'|\boldsymbol{S}_{-j}, \boldsymbol{\theta})$ and accept S_j' for all $j = 1, ..., n$ with probability

$$\Delta(S_j, S_j') = \min\left\{\frac{f(y_j|s_j', \boldsymbol{\beta})}{f(y_j|s_j, \boldsymbol{\beta})}, 1\right\}$$

Bayesian Approach: MCMC algorithm

Step 3 (Update β) Generate $\beta' \sim q(\beta, \beta')$ and accept β' with probability

$$\Delta(\beta, \beta') = \min\left\{ \frac{\prod_{i=1}^{n} f(y_i | s_i, \beta') q(\beta', \beta)}{\prod_{i=1}^{n} f(y_i | s_i, \beta) q(\beta, \beta')}, 1 \right\}$$

Step 4 (Sample $\boldsymbol{S}^*$) Generate

$$\boldsymbol{S}^* | (\boldsymbol{S}, \boldsymbol{\theta}) \sim MVN(\Sigma_{12}^T \Sigma_{11}^{-1} \boldsymbol{S},\ \ \Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12}),$$

where $\Sigma_{11} = \mathrm{Var}(\boldsymbol{S}), \Sigma_{12} = \mathrm{Cov}(\boldsymbol{S}, \boldsymbol{S}^*), \Sigma_{22} = \mathrm{Var}(\boldsymbol{S}^*)$.

Bayesian Approach: MCMC algorithm

Other efficient algorithms using spatial GLMMs have been developed.

1. Christensen and Waagepetersen (2002) suggest replacing Metropolis-Hastings step with **Langevin-Hastings updates** in the context of Poisson log-normal data.

2. Zhang (2002) combines Monte Carlo method with **EM gradient algorithm**, where spatial random effects are treated as missing data and Monte Carlo samples are used to calculate approximate conditional expectation.

More recent method

- **Limitation of MCMC:** simulation-based algorithms require tuning process and convergence analysis.
- One of more recent methods fitting spatial generalized linear mixed models in likelihood analysis framework is proposed by Bonat and Ribeiro (2015).
- Their method is based on **Laplace approximation** (Tierney and Kadane, 1986) and provides the maximized log-likelihood value, allowing for model selection and tests.

More recent method

Goal: to estimate $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \sigma^2, \tau^2, \boldsymbol{\theta})$ by maximizing the marginal likelihood function $L(\boldsymbol{\alpha})$,

$$L(\boldsymbol{\alpha}; \boldsymbol{y}(\boldsymbol{x})) = \int_{\mathbb{R}^n} f(y(\boldsymbol{x})|S(\boldsymbol{x}))f(S(\boldsymbol{x}))dS(\boldsymbol{x}).$$

Laplace approximation (Tierney and Kadane, 1986):

$$\int_{\mathbb{R}^n} \exp\{Q(\boldsymbol{u})\}d\boldsymbol{u} \approx (2\pi)^{n/2}|-Q''(\hat{\boldsymbol{u}})|^{-1/2}\exp\{Q(\hat{\boldsymbol{u}})\}, \qquad (1)$$

where $Q(\boldsymbol{u})$ is a known, unimodal, and bounded function, $\hat{\boldsymbol{u}}$ is the value for which $Q(\boldsymbol{u})$ is maximized, and $Q''(\hat{\boldsymbol{u}})$ is called Hessian.

More recent method

Assume that $f(\boldsymbol{y}(\boldsymbol{x})|S(\boldsymbol{x}))$ is an exponential family, i.e.

$$f(\boldsymbol{y}(\boldsymbol{x})|S(\boldsymbol{x})) = \exp\left\{\boldsymbol{y}(\boldsymbol{x})^T(\boldsymbol{D}\beta + S(\boldsymbol{x})) - 1^T b(\boldsymbol{D}\beta + S(\boldsymbol{x})) + 1^T c(\boldsymbol{y}(\boldsymbol{x}))\right\}$$

for some known functions $b(\cdot)$ and $c(\cdot)$.

With the usual assumption for $S(\boldsymbol{x})$, we can write the multivariate Gaussian density function of $S(\boldsymbol{x})$ as follows,

$$f(S(\boldsymbol{x}); V) = (2\pi)^{-n/2}|V|^{-1/2} \exp\left\{-\frac{1}{2}S(\boldsymbol{x})^T V^{-1} S(\boldsymbol{x})\right\},$$

with $V = \mathrm{Cov}(S(\boldsymbol{x}))$ whose (i,j) element is $C(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sigma^2\rho(\boldsymbol{x}_i - \boldsymbol{x}_j)$.

More recent method

Then, the likelihood function can be expressed as

$$L(\boldsymbol{\alpha}; \boldsymbol{y}(\boldsymbol{x})) = \int_{\mathbb{R}^n} \exp\{Q(S(\boldsymbol{x}))\} dS(\boldsymbol{x}),$$

where

$$\begin{aligned} Q(S(\boldsymbol{x})) =& \boldsymbol{y}(\boldsymbol{x})^T (\boldsymbol{D}\boldsymbol{\beta} + S(\boldsymbol{x})) - 1^T b(\boldsymbol{D}\boldsymbol{\beta} + S(\boldsymbol{x})) + 1^T c(\boldsymbol{y}(\boldsymbol{x})) \\ & - \frac{n}{2}\log(2\pi) - \frac{1}{2}\log|V| - \frac{1}{2} S(\boldsymbol{x})^T V^{-1} S(\boldsymbol{x}). \end{aligned}$$

By finding $\hat{\boldsymbol{s}} = \max Q(S(\boldsymbol{x}))$, we can approximate $L(\boldsymbol{\alpha}; \boldsymbol{y}(\boldsymbol{x}))$ using the equation (1) without calculating the integration. (Bonat and Ribeiro, 2015)

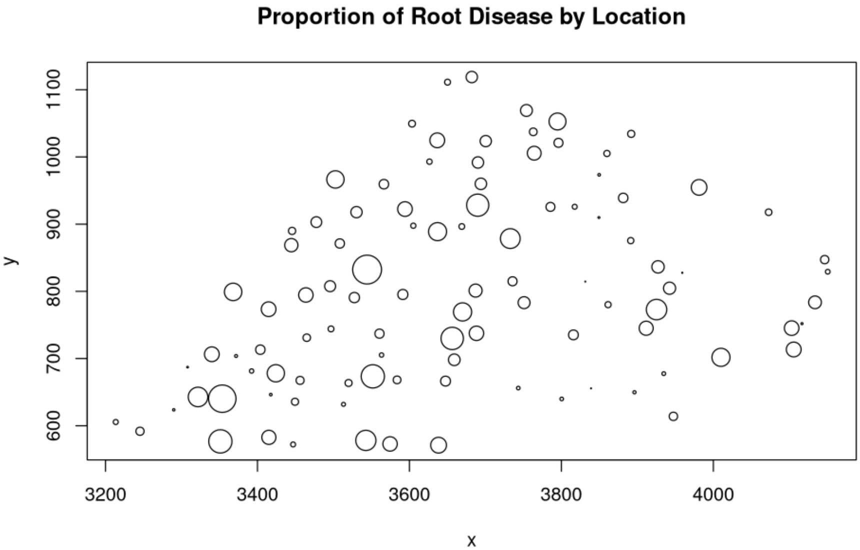
Outline

Rhizoctonia Dataset

- Rhizoctonia is a kind of fungal root rot which affects the growth of wheat and barley
- Severity can vary across a farm - mapping can provide valuable information
- Dataset contains 100 randomly selected sites from a farm in Washington. At each site, 15 plants were pulled and the total number of crown roots and number of infected crown roots were counted.

See Bonat and Ribeiro (2015) for dataset, Zhang (2002) for more details.

Map of Proportion of Root Rot



Model Fit and Prediction

Let $Y(\boldsymbol{x}_1), \dots, Y(\boldsymbol{x}_n)$ be the observed root rot counts on region B and assume the GLMM structure

$$Y(\boldsymbol{x})|S(\boldsymbol{x}) \sim \text{Binomial}(m_{\boldsymbol{x}}, p_{\boldsymbol{x}}) \quad \eta(\boldsymbol{x}) = g(\mu(\boldsymbol{x})) = \beta_0 + S(\boldsymbol{x}).$$

Here, $S(\boldsymbol{x})$ is a (stationary) mean-zero Gaussian process with spherical covariance function

$$C_S(t) = \begin{cases} \tau^2 1(t=0) + \sigma^2 \left[1 - \frac{3}{2}\frac{t}{\phi} + \frac{1}{2}\left(\frac{t}{\phi}\right)^3\right], & 0 \leq t \leq \phi \\ 0, & t > \phi \end{cases}$$

for $t = \boldsymbol{x} - \boldsymbol{x}'$ for any $\boldsymbol{x}, \boldsymbol{x}' \in B$. Our link function g is the logit link and we assume an intercept-only fixed effect. We assume the $Y(\boldsymbol{x})$ are independent conditional on $S(\boldsymbol{x})$. Paramters will be estimated using the Laplace approximation (Bonat and Ribeiro, 2015)

Model Fit and Prediction

	Estimate	Std. Error
β_0	-1.7183	0.097421
σ^2	0.10571	1.2257
ϕ	148.66	1.1187×10^{64}
τ^2	0.46676	0.43448

Table: Parameter Estimates and Standard Errors.

Note: Inferences based on the standard errors do not make sense in an analysis for such a small data set, especially covariance parameters associated with the spatial random effect. Instead, profile likelihoods can be used to quantify the uncertainty in the parameters.

Model Fit and Prediction

Let $\boldsymbol{\eta}^* = (\eta(\boldsymbol{x}_1^*), ..., \eta(\boldsymbol{x}_k^*))$ be the random variables of the link-scale mean at k unobserved spatial locations $\boldsymbol{x}^* = (\boldsymbol{x}_1^*, ..., \boldsymbol{x}_k^*)$ we wish to predict. We have

$$\mathbb{E}(\boldsymbol{\eta}) = \beta_0 1_n,\ \mathbb{E}(\boldsymbol{\eta}^*) = \beta_0 1_k$$
$$\mathrm{Cov}\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{\eta}^*\end{pmatrix} = \begin{pmatrix}\Sigma_{nn} & \Sigma_{nk}\\ \Sigma_{kn} & \Sigma_{kk}\end{pmatrix}.$$

Since the $\eta(\boldsymbol{x})$'s are normally distributed, the best heterogeneous linear predictor (known as the universal kriging predictor) has the form of the conditional mean

$$p(\boldsymbol{\eta}^*|\boldsymbol{\eta}) = \mathbb{E}(\boldsymbol{\eta}^*) + \mathrm{Cov}(\boldsymbol{\eta}^*, \boldsymbol{\eta})\,\mathrm{Cov}(\boldsymbol{\eta}, \boldsymbol{\eta})^{-1}(\boldsymbol{\eta} - \mathbb{E}(\boldsymbol{\eta}))$$

Model Fit and Prediction

Once parameter estimates are obtained, we use a universal kriging predictor on the link scale to make predictions for $g(p(\boldsymbol{x}))$ on a grid of locations $\boldsymbol{x}_1^*, ..., \boldsymbol{x}_k^* \in B$.

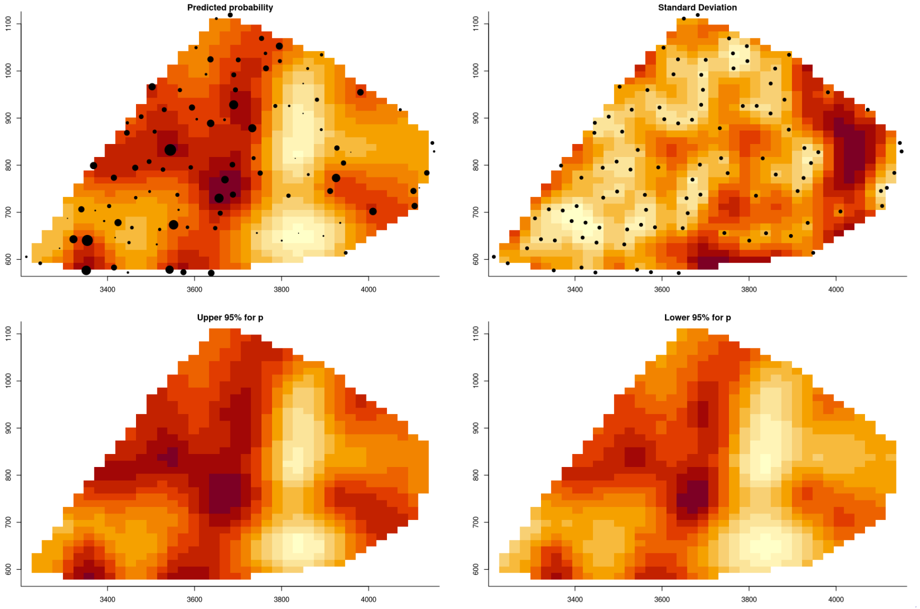
$$\tilde{\boldsymbol{\eta}}(\boldsymbol{x}^*) = \hat{\beta}_0 1_k + \hat{\Sigma}_{kn}\hat{\Sigma}_{nn}^{-1}(g(\hat{\boldsymbol{p}}) - \hat{\beta}_0 1_n)$$

where the i-th entry of $\hat{\boldsymbol{p}}$ is Y_i/m_i and the covariance matrices are estimated by plugging in the parameter estimates $\hat{\boldsymbol{\theta}} = (\hat{\tau}^2, \hat{\sigma}^2, \hat{\phi})$. Standard errors for the predictions can be derived from the estimated prediction variance

$$\hat{\Sigma}_{kk} - \hat{\Sigma}_{kn}\hat{\Sigma}_{nn}^{-1}\hat{\Sigma}_{nk}$$

We then apply the inverse logit link to get predictions on the probability scale, $p(\boldsymbol{x})$.

Prediction Maps



Take Aways

- The predicted map smooths the observed proportions while demonstrating the presence of spatial dependence.
- Prediction uncertainty is tends to be smaller closer to where we actually observed the process.
- Despite a very simple model for the mean, taking spatial dependence into account improves our understanding of the process.

References

Bonat, W.H., & Ribeiro, P.J. (2015)
Practical likelihood analysis for spatial generalized linear mixed models
Environmentrics, 27(2), 83-89.

Diggle, P.J., Tawn, J.A., & Moyeed, R.A. (1998)
Model-based geostatistics
Journal of the Royal Statistical Society: Series C (Applied Statistics), 47(3), 299-350.

Gotway, C.A., & Stroup, W.W. (1997)
A Generalized Linaer Model Approach to Spatial Data Analysis and Prediction
Journal of Agricultural, Biological, and Environmental Statistics, 2(2), 157-178.

Zhang, H. (2002)
On Estimation and Prediction for Spatial Generalized Linear Mixed Models
Biometrics, 58(1), 129-136.