Modeling spatially dependent shrub cover data in riparian zones

GMUG Meeting

December 7, 2010

Portland, OR

Bianca Eskelson

and Lisa Madsen

Forest understory vegetation

- Critical components of forest ecosystems
- Contribute to biodiversity
- Protect against erosion
- Influence nutrient cycles
- Provide forage and cover for wildlife

→ Importance of modeling understory vegetation characteristics

Abundance measures

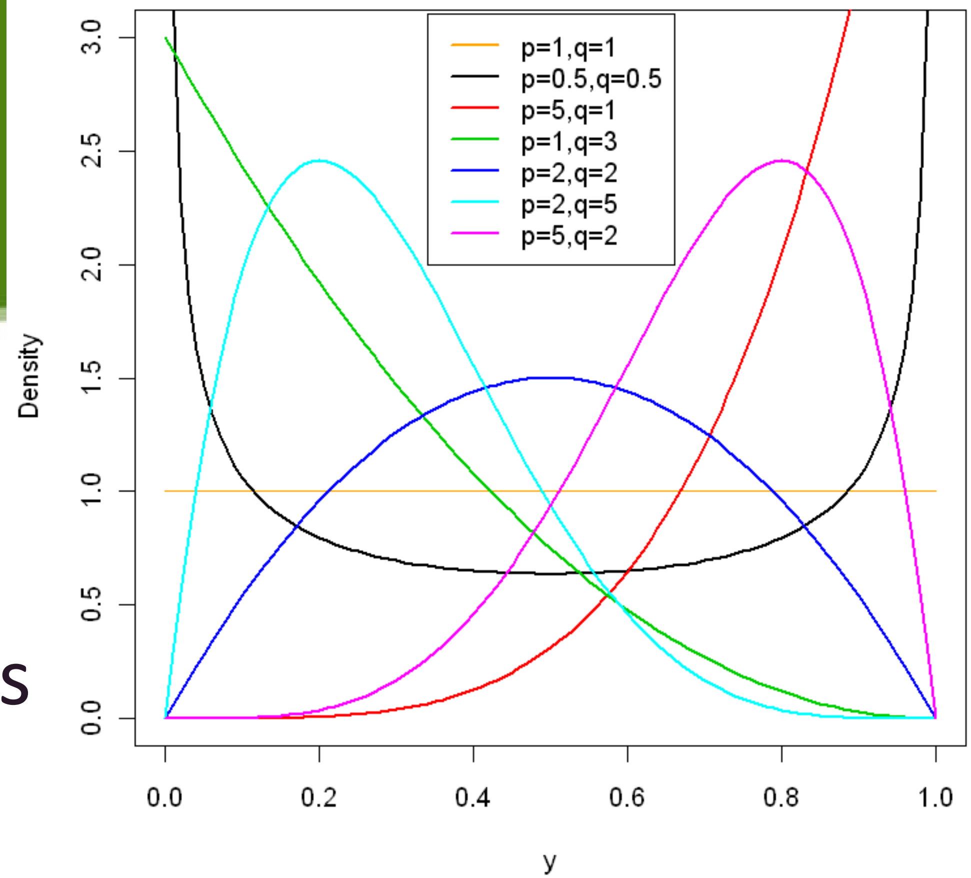
- Number of plant individuals
- Binary occurrences (presence/absence)
- Biomass per unit area
- Plant cover (e.g., percent shrub cover)
  - Prediction inherently difficult
  - Bounded between 0 and 1
  - Many zero observations & heteroscedastic error variance
  - Often subject to spatial dependence

→ Distributional features tend to be ignored in analysis

Logit Transformed Response

- Logit-transform (0,1) response to values on real line
- Standard linear regression
- Ordinary least squares (OLS)
- Generalized least squares (GLS) → account for spatial dependence

Beta Distribution



- Very flexible distribution

→ accommodates various plant cover frequency distributions

$$f\left(y;p,q\right)=\frac{\Gamma\left(p+q\right)}{\Gamma\left(p\right)\Gamma\left(q\right)}y^{p-1}\left(1-y\right)^{q-1}$$

0 < *y* < 1; shape parameters *p*, *q* > 0, and gamma function *Γ(·)*

Beta regression (BR)

- *Ferrari & Cribari-Neto (2004) proposed new parameterization where:* $\mu = \dfrac{p}{p+q}$ $\qquad \phi = p+q$

  and introduced beta regression model:

$$g\left(\mu_i\right) = x_i^T \beta_i = \eta_i$$

- *Implemented in betareg R package (Cribari-Neto and Zeileis 2010) and PROC GLIMMIX in SAS*

Copula Model

- Copula: joins univariate marginal distributions into multivariate distribution function
- Multivariate Gaussian copula generalizes multivariate normal dependence structure to non-normal marginals
- A joint distribution function is

$$C\left(y;\Sigma\right)=\Phi_{\Sigma}\left[\Phi^{-1}\left\{F_1\left(y_1\right)\right\},\ldots,\Phi^{-1}\left\{F_n\left(y_n\right)\right\}\right]$$

Φ standard normal cdf

Φ_{Σ} multivariate normal cdf with correlation matrix Σ

Gaussian Copula Joint Density

- Differentiating the distribution funtion yields the joint density function:

$$c(y;\Sigma) = |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{z}^T\left(\Sigma^{-1} - \boldsymbol{I}_n\right)\mathbf{z}\right)\prod_{i=1}^{n} f_i(y_i)$$

$$\mathbf{z} = \left[\Phi^{-1}\left\{F_1(y_1)\right\}, \ldots, \Phi^{-1}\left\{F_n(y_n)\right\}\right]^T$$

$\boldsymbol{I}_n$ denotes the *n* x *n* identity matrix

f_i is the marginal density of y_i

Σ determines the dependence structure

Spatial Gaussian Copula

- Spatial correlation matrix with exponential 'decay' parameter θ:

$$\Sigma_{ij}(\theta) = \begin{cases} \exp\left(-h_{ij}\theta\right), & i \neq j \\ 1, & i = j \end{cases}$$

- Spatial Gaussian copula brings non-normal distributions into Gaussian geostatistical framework (Madsen 2009)

- Obtain maximum likelihood estimates by numerically maximizing the log of expected likelihood with respect to $\boldsymbol{\beta}$, ϕ, and θ

Objectives

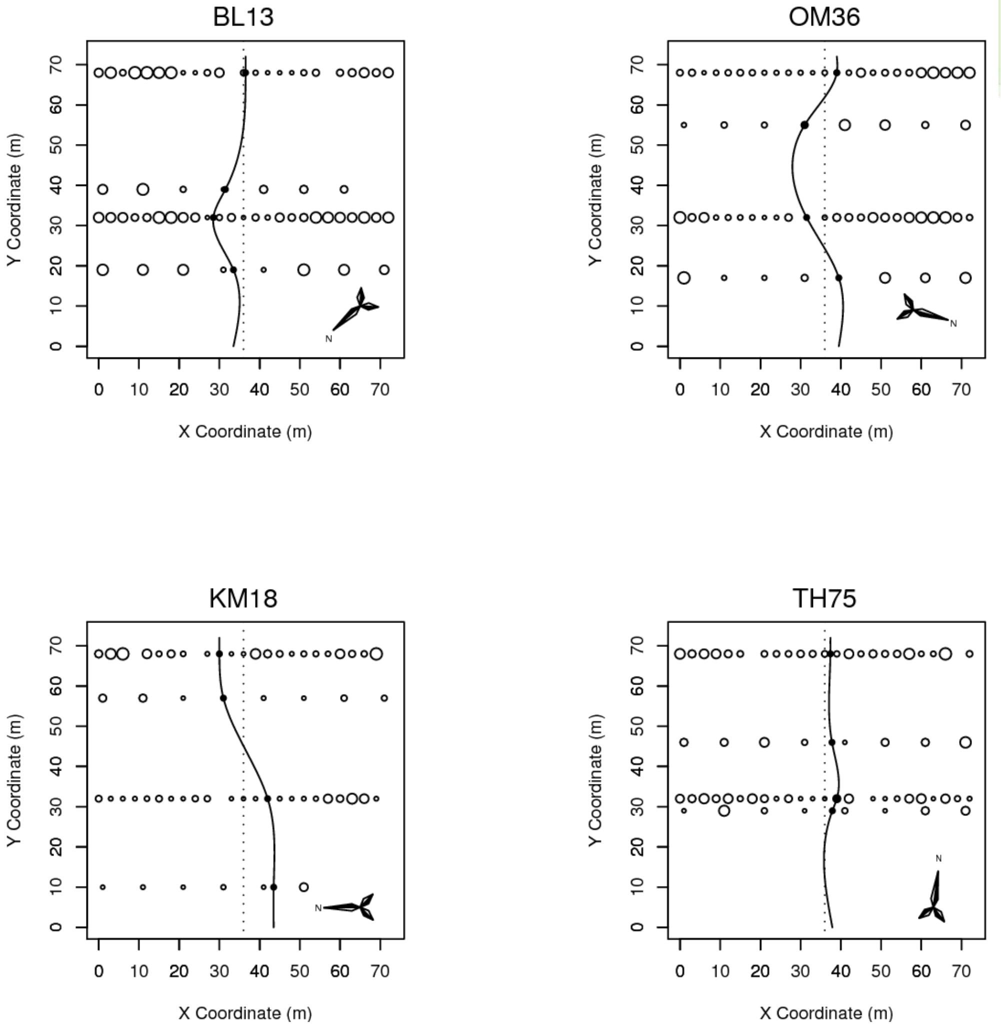
- *Case Study*: Model % shrub cover in riparian forests as function of topographic conditions and overstory vegetation characteristics using five modeling approaches:

  OLS, GLS, BR, BRdep, and COP

- *Simulation Study*: Evaluate the performance of five modeling approaches in terms of their parameter estimates

Case Study – Data

- 4 headwater streams in the Oregon Coast Range (Cissel et al. 2006, Marquardt 2010)
- Percent shrub cover
- Visually determined (nearest 5%) on 1 m x 1 m plots along transects (n=248)
- Distance to stream (DTS in m)
- Height above stream (HAS in m)
- Leaf area index (LAI)
- % slope, aspect



Case Study – Results

- Model with smallest Bayesian Information Criterion included DTS & LAI as covariates

- BR models had poor explanatory power (pseudo-$R^2 \leq 0.34$)

- *OLS & GLS models*: MSPE largest, negative bias

- *BR, BRdep & COP models*: MSPE slightly smaller, unbiased

Simulation Study – Data

- 500 data sets of size n=248 to mimic observed shrub data
- Simulated response from beta distribution with

$$\mu_i = \frac{1}{1+\exp\left(-\left(\beta_0 + \beta_1 DTS_i + \beta_2 LAI_i\right)\right)}$$

and $(\beta_0, \beta_1, \beta_2, \phi) = (-0.34, 0.037, -0.24, 2.6)$

- Spatial locations, DTS, and LAI agree with actual data from case study
- Spatial dependence in simulated response:
  - Simulate spatially dependent standard normal random variables Z
  - Covariance matrix based on exponential model with θ ranging from 0.01 (strong spatial dependence) to infinity (no spatial dependence)

$$\sum_{ij}(\theta) = \begin{cases} \exp\left(-h_{ij}\theta\right), & i \neq j \\ 1, & i = j \end{cases}$$

  - Transform Z's to beta response Y

Simulation Study – Methods

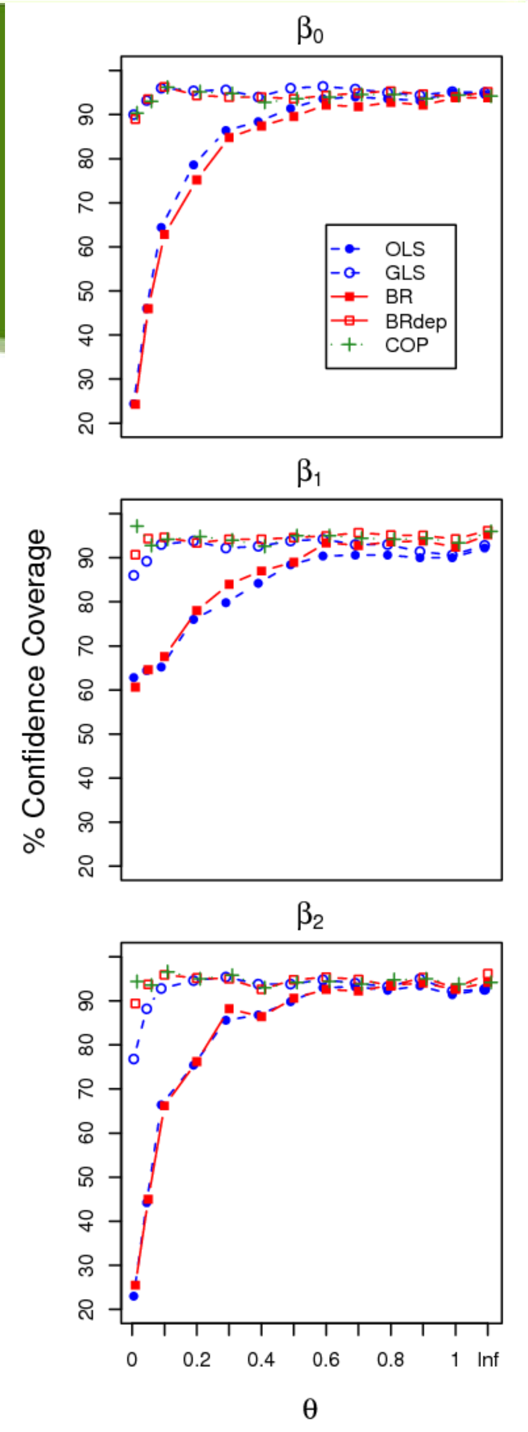
- Fit OLS, GLS, BR, BRdep, and COP models to 500 simulated data sets across range of θ values
- Calculate confidence coverage (CC) of parameters β_0, β_1, β_2

NOTE: OLS and GLS parameters can be expressed in terms of β_0, β_1, β_2 of the BR, BRdep, and COP models using Equation 15 in Espinheira et al. (2008)

Results – Simulation Study (cont'd)

- *OLS & BR*: poor CC for all β's when spatial dependence is strong (small θ)

- *GLS, BRdep & COP*:
  - 95% confidence coverage (CC) for all β's for $\theta \geq 0.1$
  - 77-95% CC for $\theta < 0.1$ with GLS having smallest CC



Conclusions

- Model fit is poor due to scale issues
- OLS and GLS resulted in biased model predictions → not recommended for modeling % shrub cover
- BR, BRdep, and COP provided unbiased predictions
- OLS & BR should not be used in the presence of spatial dependence
- When spatial dependence is strong, GLS, BRdep, and COP result in confidence coverage < 95%, with GLS performing worst
- BR and COP models should be extended to account for zero-inflation

Possible Future Applications

- Copula model not restricted to shrub cover
- Model % canopy cover
- Model crown ratio → currently working on a copula model for modeling crown ratio and tree height simultaneously

Acknowledgments

- Theresa Marquardt for help with the data
- Data collection was funded by the USDA Forest Service, Dr. Paul Anderson
- USGS for funding
- Co-authors Joan Hagar and Temesgen Hailemariam for comments and support

References

Cissel, J.H., Anderson, P.D., Olson, D., Puettmann, K.P., Berryman, S., Chan, S.S., Thompson, C. 2006. BLM Density Management and Riparian Buffer Study: Establishment Report and Study Plan, U.S. Geological Survey, Scientific Investigations Report 2006–5087, 151 p.

Cribari-Neto, F. and A. Zeileis. Beta regression in R. J. Stat. Softw. 34(2):1-24.

Espinheira, P.L., Ferrari, S.L.P., and F. Cribari-Neto. 2008. On beta regression residuals. J. Appl. Stat. 35(4):407-419.

Ferrari, S.L.P., and F. Cribari-Neto. 2004. Beta regression for modelling rates and proportions. J. Appl. Stat. 31(7):799-815.

Madsen, L. 2009. Maximum likelihood estimation of regression parameters with spatially discrete data. J. Agr. Biol. Environ. Stat. 14:375-391.

Marquardt, T. 2010. Accuracy and suitability of several stand sampling methods in riparian zones. MS thesis. Oregon State University. 77p.

Questions?

For more details see:

Eskelson, B.N.I., Madsen, L., Hagar, J., and H. Temesgen. *In press*. Estimating riparian understory vegetation cover with beta regression and copula models. Forest Science XX:xxx-xxx.