Composition Models for Benthic Invertebrate Data

Devin Johnson, Jennifer Hoeting, LeRoy Poff,
and
Brian Bledsoe

Colorado State University

Research supported by

- Environmental Protection Agency grants R-82909501 and R-82863601
- National Science Foundation grant DEB-0091961.

Species Composition Models

- Direct observation of species, or functional group, abundance has long been a key indicator of ecological health

- Two items are critical for development of biological monitoring programs:
  - Understanding of how environmental variables, at different scales, affect abundance and composition of species or functional groups
  - Which species or functional/life history traits are most beneficial to examine

Single Trait State-Space Model

Once collected at a stream site, observed species are categorized into different levels of a single trait (e.g. drift propensity, feeding type, or body shape).

In addition, a suite of environmental covariates, at both local and watershed scales, are also observed.

notation:

- i Site index ($i = 1, \ldots, S$)
- j Trait level index ($j = 1, \ldots, J$)
- k Local scale environmental variable index ($k = 1, \ldots, L$)
- l Watershed scale environmental variable index ($l = 1, \ldots, W$)

Single Trait State-Space Model

Response:

$$Y_{ij}|\lambda_{ij} \sim \text{indep. Poisson}(\lambda_{ij})$$

$$\log \lambda_{ij} = \theta_{ij} + X_i' \beta_j$$

where,

- Y_{ij} is the number of organisms with trait level j at site i
- X_i is an $L + W$ vector of combined scale covariates
- $\theta_i = (\theta_{i1}, \ldots, \theta_{iJ}) \;\sim\; MVN\left(\mu_\theta,\; \mathbf{T}_\theta^{-1}\right)$ for $i = 1, \ldots, S$
  (Space-time correlation could be modeled)
- Rate composition: $\lambda_{ij} / \sum_{j=1}^{J} \lambda_{ij} \;\sim\;$ logistic normal

Single Trait State-Space Model

Covariates: Let $X_i = (X_i^{(L)},\ X_i^{(W)})$, where $X_i^{(L)}$ is an L vector of local environmental covariates and $X_i^{(W)}$ is a W vector of watershed covariates. Then, we consider the covariate model:

$$X_i \ \sim \ MVN\left(\mu_X,\ \mathbf{T}_X^{-1}\right)$$

or, equivalently,

$$X_i^{(L)} | X_i^{(W)} \ \sim \ MVN\left(\gamma_0 + \gamma' X_i^{(W)},\ \mathbf{T}_{L|W}^{-1}\right)$$

and

$$X_i^{(W)} \ \sim \ MVN\left(\mu_W,\ \mathbf{T}_W^{-1}\right),$$

where,

$$\gamma = \text{Var}\left(X_i^{(W)}\right)^{-1} \text{Cov}\left(X_i^{(L)},\ X_i^{(W)}\right)$$

Number of Reproductive Generations per Year

- Data collected as part of 1994 and 1995 Colorado and Oregon Regional EMAP studies

- **Response**:

  Each species is classified according to one of three categories

  1. Semi-volitine ($<$ 1 reproductive generation per year)
  2. Univolitine (1 reproductive generation per year)
  3. Multi-volitine (more than 1 reproductive generation per year)

Number of Reproductive Generations per Year

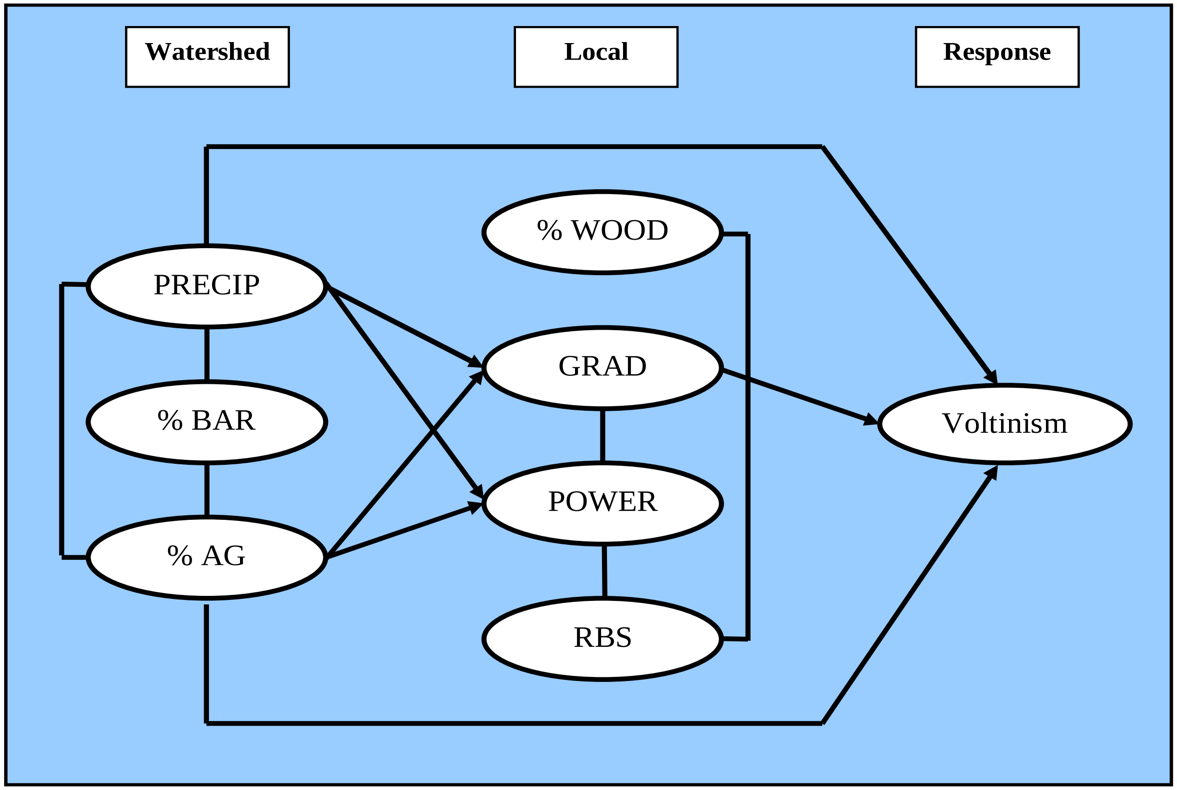
- **Local covariates**:
  - % WOOD: % of substrate composed of wood
  - GRAD: Gradation coefficient
  - POWER: Surrogate for stream power
  - RBS: Relative bed stability

- **Watershed covariates**:
  - PRECIP: Average number of inches over the basin
  - % BAR: % barren land in basin
  - % AG: % agricultural pasture in basin

- Using vague prior information, parameters are estimated using a Gibbs MCMC sample

Number of Reproductive Generations per Year

Graphical Model



Edge inclusion determined by credible intervals for the elements of β, γ, $\mathbf{T}_W$, and $\mathbf{T}_{L|W}$

Continuing Research

We are currently investigating the following research topics:

- Bayesian estimation and model determination
  - Stochastic search over possible graphical models (edge addition/deletion)
  - Efficient parameterizations and algorithms for MCMC estimation
- Additional parameterizations/models
  - Addition of spatial correlation structure
  - Multiple compositions
  - Modeling discrete and continuous covariates (Conditional Gaussian distribution)