Scalable and robust regression models for continuous proportional data

Changwoo Lee

Postdoctoral Associate, Department of Statistical Science, Duke University

Apr 2025

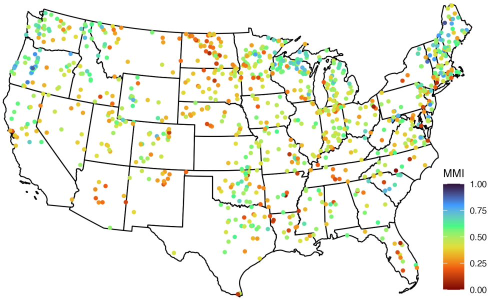
Joint work with B. Dahl (Duke), O. Ovaskainen (U. Jyväskylä), and D. Dunson (Duke)

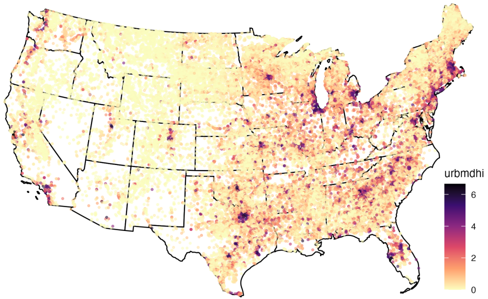
Continuous proportional data

Response Y in the unit interval $[0, 1]$, non-count based proportions

- (Medical imaging) Percentage of tissue area in mammogram [Peplonska et al., 2012]
- (Econometrics) Central-bank independence index [Berggren et al., 2014]
- (Political science) Suffrage (voting rights) index [Kubinec, 2023]
- Especially common in ecology:
  - Percent cover measurements [de Vargas Ribeiro et al., 2022][Korhonen et al., 2024][van Strien et al., 2024]
  - Diversity index [Lindholm et al., 2021][Bharti et al., 2023][Rolls et al., 2023][Qiao et al., 2025]
  - [Warton and Hui, 2011]: $\approx$14% of ecology papers involves non-count based proportions.

Benthic macroinvertebrate multimetric index (MMI)





- Higher MMI indicates a healthier/diverse lake benthic macroinvertebrate community
- (Left): MMI of 949 lakes from 2017 NLA survey [U.S. Environmental Protection Agency, 2022]
- (Right): Lake watershed covariate of 50k+ lakes from LakeCat data [Hill et al., 2018]
  - medium/high urban land cover, soil erodibility factor, coniferous forest cover, ...

Modeling continuous proportional data

- Approach 1: **Transform** $Y \mapsto f(Y) \in \mathbb{R}$ (e.g. logit) and fit normal linear model.
  - Interpretation based on $E(f(Y) \mid X)$, not $E(Y \mid X)$
  - Problematic for boundary values.
- Approach 2: **Beta regression** [Ferrari and Cribari-Neto, 2004, Cribari-Neto and Zeileis, 2010]

$$Y_i \mid \boldsymbol{x}_i \overset{\text{ind}}{\sim} \text{Beta}(\mu_i, \phi), g(\mu_i) = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta},$$

$$p_{\text{beta}}(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma(\phi - \mu\phi)} y^{\mu\phi - 1}(1-y)^{\phi - \mu\phi - 1}, \quad 0 < y < 1$$

  - Mean & precision parameterization, $E(Y) = \mu$, $\text{var}(Y) = \mu(1-\mu)/(1+\phi)$
  - Familiar GLM-like structure, $g(E(y \mid \boldsymbol{x})) = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}$, with a choice of link function g.

Limitations of beta regression

- **Non-robustness**: sensitive to violation of beta response assumption.
  - Beta does not belong to a natural exponential family
  - Does not belong to GLM, μ and ϕ not orthogonal to each other [Ferrari and Cribari-Neto, 2004]
- **Computational challenges**: Bayesian inference in hierarchical settings.
  - Mixed models, longitudinal and spatial models: generic methods (e.g. `Stan`) may suffer
  - Existing scalable methods cannot be easily applied, unless relying on approximation
- **Handling boundary values**: cannot handle exact 0s and 1s
  - Preprocess ("nudge") the data to lie between open interval $(0, 1)$ [Smithson and Verkuilen, 2006]
  - Results are often sensitive [Kosmidis and Zeileis, 2024], not desirable

Limitations of beta regression

- **Non-robustness**: sensitive to violation of beta response assumption.
  - Beta does not belong to a natural exponential family
  - Does not belong to GLM, μ and ϕ not orthogonal to each other [Ferrari and Cribari-Neto, 2004]
- **Computational challenges**: Bayesian inference in hierarchical settings.
  - Mixed models, longitudinal and spatial models: generic methods (e.g. `Stan`) may suffer
  - Existing scalable methods cannot be easily applied, unless relying on approximation
- **Handling boundary values**: cannot handle exact 0s and 1s
  - Preprocess ("nudge") the data to lie between open interval $(0, 1)$ [Smithson and Verkuilen, 2006]
  - Results are often sensitive [Kosmidis and Zeileis, 2024], not desirable

Limitations of beta regression

- **Non-robustness**: sensitive to violation of beta response assumption.
  - Beta does not belong to a natural exponential family
  - Does not belong to GLM, μ and ϕ not orthogonal to each other [Ferrari and Cribari-Neto, 2004]
- **Computational challenges**: Bayesian inference in hierarchical settings.
  - Mixed models, longitudinal and spatial models: generic methods (e.g. `Stan`) may suffer
  - Existing scalable methods cannot be easily applied, unless relying on approximation
- **Handling boundary values**: cannot handle exact 0s and 1s
  - Preprocess ("nudge") the data to lie between open interval $(0, 1)$ [Smithson and Verkuilen, 2006]
  - Results are often sensitive [Kosmidis and Zeileis, 2024], not desirable

Summary of contribution

- **Cobin regression**: continuous binomial (cobin) regression model
  - A proper GLM approach based on **exponential dispersion model** [Jørgensen, 1987]
  - Inherits attractive properties of GLM, including robustness of $\hat{\beta}$ to model misspecificaton
- **Micobin regression**: based on dispersion mixtures of cobin distributions
  - More **flexible** & **robust** family of distribution (cf. t dist as scale mixture of normal)
  - **Can handle exact 0s and 1s**, different from structural zeros [Blasco-Moreno et al., 2019]
- We introduce **Kolmogorov-Gamma augmentation** for Bayesian computation
  - Converts cobin/micobin likelihood into **conditionally normal likelihood**
  - Seamless integration with scalable methods for latent Gaussian models

Summary of contribution

- **Cobin regression**: continuous binomial (cobin) regression model
  - A proper GLM approach based on **exponential dispersion model** [Jørgensen, 1987]
  - Inherits attractive properties of GLM, including robustness of $\hat{\beta}$ to model misspecificaton
- **Micobin regression**: based on dispersion mixtures of cobin distributions
  - More **flexible** & **robust** family of distribution (cf. t dist as scale mixture of normal)
  - **Can handle exact 0s and 1s**, different from structural zeros [Blasco-Moreno et al., 2019]
- We introduce **Kolmogorov-Gamma augmentation** for Bayesian computation
  - Converts cobin/micobin likelihood into **conditionally normal likelihood**
  - Seamless integration with scalable methods for latent Gaussian models

Summary of contribution

- **Cobin regression**: continuous binomial (cobin) regression model
  - A proper GLM approach based on **exponential dispersion model** [Jørgensen, 1987]
  - Inherits attractive properties of GLM, including robustness of $\hat{\beta}$ to model misspecificaton
- **Micobin regression**: based on dispersion mixtures of cobin distributions
  - More **flexible** & **robust** family of distribution (cf. t dist as scale mixture of normal)
  - **Can handle exact 0s and 1s**, different from structural zeros [Blasco-Moreno et al., 2019]
- We introduce **Kolmogorov-Gamma augmentation** for Bayesian computation
  - Converts cobin/micobin likelihood into **conditionally normal likelihood**
  - Seamless integration with scalable methods for latent Gaussian models

Exponential dispersion model (1)

- Exponential dispersion model [Jørgensen, 1987] offers a principled way to add dispersion parameter λ^{-1} that controls variance to one-parameter natural exponential family

	exponential tilting by θ $\longrightarrow$	λ-fold convolution and scale by λ^{-1} $\longrightarrow$
$H_1 \sim N(0,1)$ $h_1(y) = (2\pi)^{-1/2}\exp(-y^2/2)$ $y \in \mathbb{R}$	$Y \sim N(\theta, 1)$ $\exp(\theta y - B_1(\theta))h_1(y)$ $\theta \in \mathbb{R}$	$Y \sim N(\theta, \lambda^{-1})$ $\exp(\lambda\theta y - \lambda B_1(\theta))h_1(y,\lambda)$ $\lambda > 0$

$B_1(\theta) = \log E(e^{\theta H_1}) = \frac{\theta^2}{2}$, $\quad h_1(y,\lambda) =$ density of $\frac{1}{\lambda}\sum_{l=1}^{\lambda} H_1^{(l)} = \frac{\exp(-\lambda y^2/2)}{\sqrt{2\pi/\lambda}}$

Exponential dispersion model (2)

- Exponential dispersion model [Jørgensen, 1987] offers a principled way to add dispersion parameter λ^{-1} that controls variance to one-parameter natural exponential family

	exponential tilting by θ $\longrightarrow$	λ-fold convolution and scale by λ^{-1} $\longrightarrow$
$H_2 \sim \text{Exp}(1)$	$Y \sim \text{Exp}(1-\theta)$	$Y \sim \text{Gamma}(\lambda, \lambda(1-\theta))$
$h_2(y) = \exp(-y)$	$\exp(\theta y - B_2(\theta))h_2(y)$	$\exp(\lambda\theta y - \lambda B_2(\theta))h_2(y,\lambda)$
$y \geq 0$	$\theta < 1$	$\lambda > 0$

$B_2(\theta) = \log E(e^{\theta H_2}) = -\log(1-\theta)$, $h_2(y,\lambda) =$ density of $\frac{1}{\lambda}\sum_{l=1}^{\lambda} H_2^{(l)} = \frac{\lambda^\lambda y^{\lambda-1}\exp(-\lambda y)}{\Gamma(\lambda)}$

Exponential dispersion model (3)

- Exponential dispersion model [Jørgensen, 1987] offers a principled way to add dispersion parameter λ^{-1} that controls variance to one-parameter natural exponential family

	exponential tilting by θ $\longrightarrow$	λ-fold convolution and scale by λ^{-1} $\longrightarrow$
$H_3 \sim \mathrm{InvGamma}(1/2, 1/2)$	$Y \sim \mathrm{InvGau}((-2\theta)^{-1/2}, 1)$	$Y \sim \mathrm{InvGau}((-2\theta)^{-1/2}, \lambda)$
$h_3(y) = (2\pi y^3)^{-1/2} \exp(-\frac{1}{2y})$	$\exp(\theta y - B_3(\theta))h_3(y)$	$\exp(\lambda\theta y - \lambda B_3(\theta))h_3(y, \lambda)$
$y > 0$	$\theta < 0$	$\lambda > 0$

$B_3(\theta) = \log E(e^{\theta H_3}) = -(-2\theta)^{1/2}$, $\quad h_3(y, \lambda) =$ density of $\frac{1}{\lambda}\sum_{l=1}^{\lambda} H_3^{(l)} = \frac{\lambda^{1/2}\exp(-\lambda/(2y))}{\sqrt{2\pi y^3}}$

Exponential dispersion model (4)

	exponential tilting by θ $\longrightarrow$	λ-fold convolution and scale by λ^{-1} $\longrightarrow$
$H \sim \mathrm{Unif}(0,1)$	$Y \sim \mathrm{cobin}(\theta, 1)$	$Y \sim \mathrm{cobin}(\theta, \lambda^{-1})$
$h(y) = 1$	$\exp(\theta y - B(\theta))h(y)$	$\exp(\lambda\theta y - \lambda B(\theta))h(y,\lambda)$
$0 \le y \le 1$	$\theta \in \mathbb{R}$	$\lambda \in \mathbb{N}$

$B(\theta) = \log E(e^{\theta H}) = \log\left(\frac{e^\theta - 1}{\theta}\right)$, $h(y,\lambda) =$ density of $\frac{1}{\lambda}\sum_{l=1}^{\lambda} H^{(l)}$ [Bates, 1955]

$$h(y,\lambda) = \frac{\lambda}{(\lambda-1)!}\sum_{k=0}^{\lambda}(-1)^k\binom{\lambda}{k}\{\max(\lambda y - k, 0)\}^{\lambda-1}$$

Continuous binomial distribution

Definition 1. (cobin)

We say $Y \sim \mathrm{cobin}(\theta, \lambda^{-1})$, natural param. $\theta \in \mathbb{R}$, dispersion $\lambda^{-1} \in \{1, 1/2, \dots\}$ if

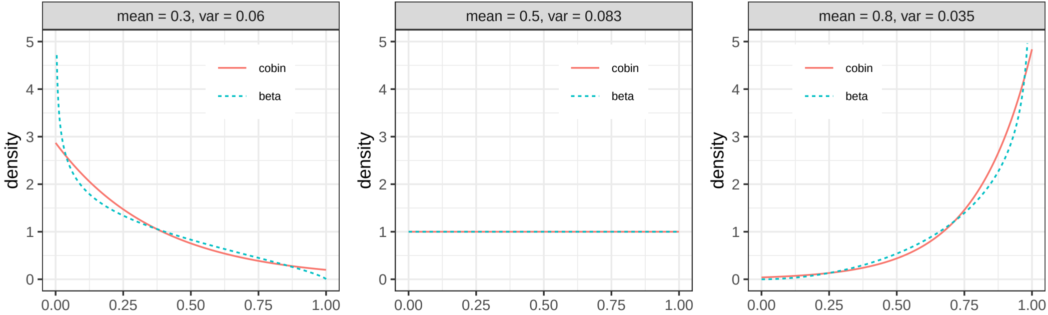
$$p_{\mathrm{cobin}}(y; \theta, \lambda^{-1}) = h(y, \lambda) \exp\left[\lambda\{\theta y - B(\theta)\}\right] = h(y, \lambda) \frac{e^{\lambda\theta y}}{\{(e^{\theta} - 1)/\theta\}^{\lambda}}, \quad 0 \le y \le 1$$

- Exponential dispersion model [Jørgensen, 1987]; $E(Y) = B'(\theta)$, $\mathrm{var}(Y) = \lambda^{-1} B''(\theta)$.
- λ must be an integer to be a valid distribution
- Name motivated from continuous Bernoulli [Loaiza-Ganem and Cunningham, 2019] ($\lambda = 1$ case),

$$Y_1, \dots, Y_\lambda \overset{\text{iid}}{\sim} \mathrm{cobin}(\theta, 1) \implies \tfrac{1}{\lambda} \textstyle\sum_{l=1}^{\lambda} Y_l \sim \mathrm{cobin}(\theta, \lambda^{-1})$$

Cobin vs beta

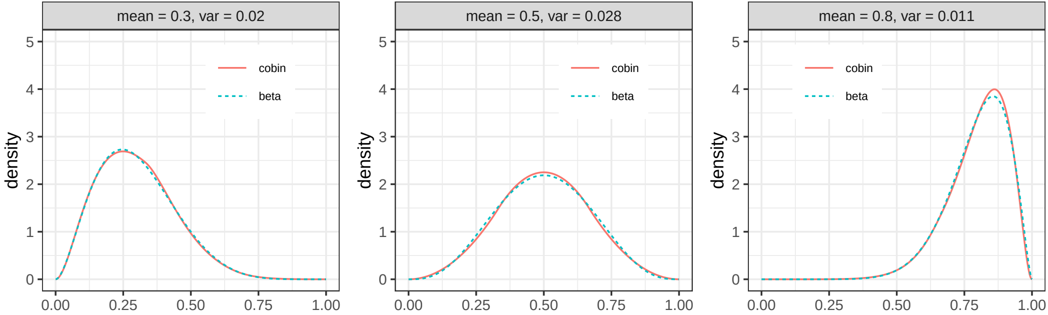
Cobin with $\lambda = 1$; comparison with beta under the same mean and variance



Support: closed interval $[0, 1]$ for cobin with $\lambda = 1$ (continuous Bernoulli)

Cobin vs beta

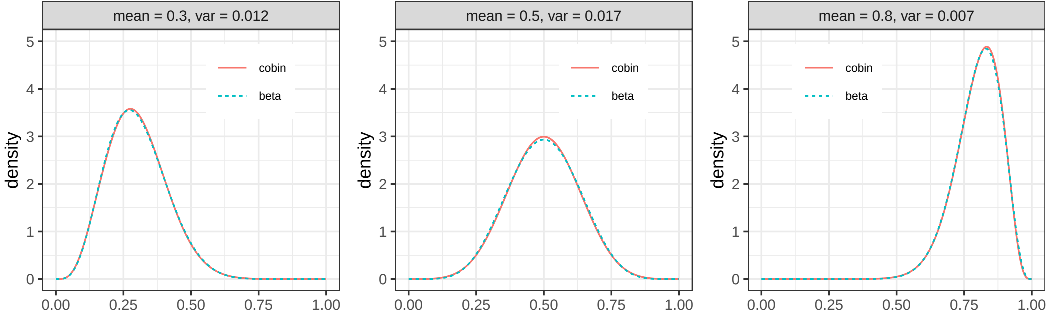
Cobin with $\lambda = 3$; comparison with beta under the same mean and variance



Support: open interval $(0, 1)$ for cobin with $\lambda > 1$

Cobin vs beta

Cobin with $\lambda = 5$; comparison with beta under the same mean and variance



Support: open interval $(0, 1)$ for cobin with $\lambda > 1$

Cobin regression model

- With a choice of link function $g : (0,1) \to \mathbb{R}$,

$$Y_i \mid \theta_i, \lambda \overset{\text{ind}}{\sim} \text{cobin}(\theta_i, \lambda^{-1}), \quad \theta_i = (B')^{-1}\{g^{-1}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})\}, \quad i = 1, \ldots, n, \tag{1}$$

implies $E(Y_i \mid \boldsymbol{x}_i) = g^{-1}(\eta_i) = g^{-1}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})$. **Proper GLM** [Nelder and Wedderburn, 1972]

- Likelihood equations (score function):

$$\frac{\partial}{\partial \beta_j}\sum_{i=1}^n \log p_{\text{cobin}}(y_i; \theta_i, \lambda^{-1}) = \lambda \sum_{i=1}^n \frac{(y_i - \mu_i)x_{ij}}{B''(\theta_i)}\frac{\partial \mu_i}{\partial \eta_i} = 0, \quad j = 1, \ldots, p \tag{2}$$

cf. beta score function: $\phi \sum_{i=1}^n [\log \frac{y_i}{1-y_i} - \Psi(\mu_i\phi) + \Psi(\phi - \mu_i\phi)] x_{ij} \frac{\partial \mu_i}{\partial \eta_i}$. ($\Psi$: digamma ft).

Proposition (consistency of cobin MLE under potential misspecification)

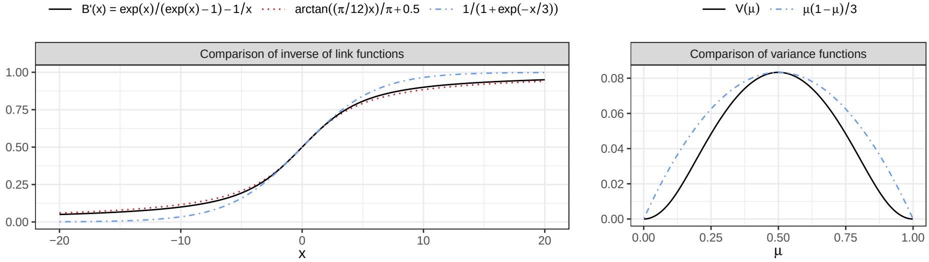
If the mean structure $E(y_i \mid x_i) = g^{-1}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})$ is correctly specified, the solution $\hat{\beta}$ of the cobin regression likelihood equations (2) is **consistent** [Gourieroux et al., 1984]

Cobin regression with canonical link function

- Canonical link function: $g_{\text{cobit}} = (B')^{-1}$ (“cobit”) that leads to $\theta_i = \eta_i = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}$
  - Cobin regression likelihood becomes

$$\prod_{i=1}^{n} p_{\text{cobin}}(y_i; \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}, \lambda^{-1}) = \prod_{i=1}^{n} h(y_i, \lambda)\frac{e^{\lambda y_i \eta_i}}{\{(e^{\eta_i}-1)/\eta_i\}^{\lambda}}$$

  - $g_{\text{cobit}}^{-1}(\eta) = e^{\eta}/(e^{\eta}-1) - 1/\eta$ is similar to Cauchy c.d.f.



Micobin: dispersion mixtures of cobin distributions

- Limitations of cobin:
  - λ must be an integer to be a valid distribution $\implies$ **limited flexibility**
  - **Unnatural to model dispersion** λ with a separate set of covariates due to discreteness
  - Unless $\lambda = 1$, cobin is supported on open interval $(0, 1)$, **cannot handle exact 0s and 1s**

Definition 2. (micobin)

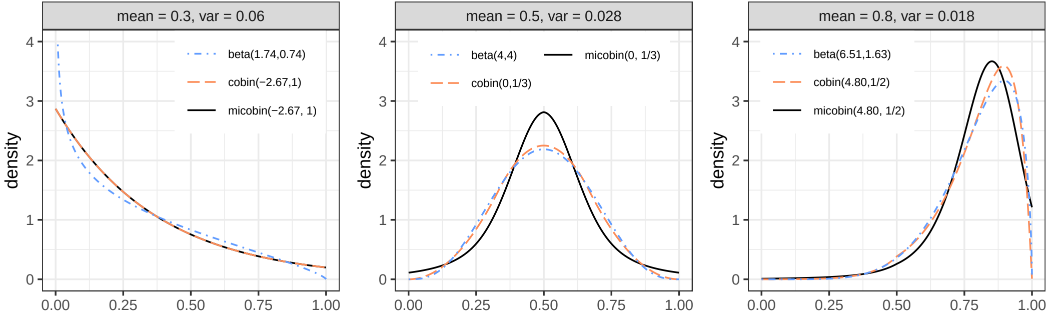
We say $Y \sim \mathrm{micobin}(\theta, \psi)$, natural param. $\theta \in \mathbb{R}$, dispersion $\psi \in (0, 1)$ if

$$Y \mid \lambda \sim \mathrm{cobin}(\theta, \lambda^{-1}), \quad (\lambda - 1) \sim \mathrm{negbin}(2, \psi)$$

- Dispersion $\psi \in (0, 1)$, mean structure preserved $E(Y) = B'(\theta)$ by $\psi-$mixture.
- $(\lambda - 1) \sim \mathrm{negbin}(2, \psi)$ leads to $\mathrm{var}(Y) = \psi B''(\theta)$. (cf. $\mathrm{var}(Y) = \lambda^{-1} B''(\theta)$ for cobin)
- Localizing dispersion ($\lambda \to \lambda_i$), general approach for **robustification** [Wang and Blei, 2018]

Micobin: dispersion mixtures of cobin distributions

Comparison of beta, cobin, and micobin with the same mean and variance.



Support of micobin is a closed interval $[0, 1]$ for any θ, ψ.

Micobin regression and extensions

- Micobin regression with cobit link $E(Y_i \mid \boldsymbol{x}_i) = g_{\text{cobit}}^{-1}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})$:

$$Y_i \mid \boldsymbol{x}_i \sim \operatorname{micobin}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}, \psi) \iff Y_i \mid \boldsymbol{x}_i, \lambda_i \sim \operatorname{cobin}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}, \lambda_i^{-1}), \quad (\lambda_i - 1) \sim \operatorname{negbin}(2, \psi)$$

- Can be extended to accommodate dispersion covariate: $\operatorname{logit}(\psi_i) = \boldsymbol{d}_i^{\mathrm{T}}\gamma_i$
- Further hierarchical extensions (with cobit link):
  - **Random intercept model**

$$Y_{ij} \sim \operatorname{micobin}(\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + u_i, \psi), \quad u_i \overset{\text{iid}}{\sim} N(0, \sigma_u^2)$$

  - **Spatial regression model** with spatially indexed data $(y(s_i), \boldsymbol{x}(s_i))$:

$$\begin{aligned} Y(s_i) &\sim \operatorname{micobin}(\eta_i, \psi) \\ \eta_i &= \boldsymbol{x}(s_i)^{\mathrm{T}}\boldsymbol{\beta} + u(s_i), \quad u(\cdot) \sim \text{mean zero GP}. \end{aligned}$$

Data augmentation

- Recall the cobin regression likelihood under the canonical link $\theta_i = \eta_i = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}$

$$p(y_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}, \lambda) = p_{\text{cobin}}(y_i; \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}, \lambda^{-1}) = h(y_i, \lambda)\frac{e^{\lambda y_i \eta_i}}{\{(e^{\eta_i} - 1)/\eta_i\}^{\lambda}}$$

- Not a familiar expression in terms of η_i
- For latent Gaussian models, we desire a log-likelihood that is a quadratic function in η_i
- **Data augmentation**: introduce (missing) auxiliary data $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_n)$ such that
  - Retain original model upon marginalization: $\int p(y_i, \kappa_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}, \lambda)\mathrm{d}\kappa_i = p(y_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}, \lambda)$
  - Conditional distributions $p(y_i \mid \kappa_i, \boldsymbol{x}_i, \boldsymbol{\beta}, \lambda)$, $p(\kappa_i \mid y_i, \boldsymbol{x}_i, \boldsymbol{\beta}, \lambda)$ are easy to work with.

Data augmentation

- Recall the cobin regression likelihood under the canonical link $\theta_i = \eta_i = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}$

$$p(y_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}, \lambda) = p_{\text{cobin}}(y_i; \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}, \lambda^{-1}) = h(y_i, \lambda)\frac{e^{\lambda y_i \eta_i}}{\{(e^{\eta_i} - 1)/\eta_i\}^{\lambda}}$$

- Not a familiar expression in terms of η_i
- For latent Gaussian models, we desire a log-likelihood that is a quadratic function in η_i
- **Data augmentation**: introduce (missing) auxiliary data $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_n)$ such that
  - Retain original model upon marginalization: $\int p(y_i, \kappa_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}, \lambda)\mathrm{d}\kappa_i = p(y_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}, \lambda)$
  - Conditional distributions $p(y_i \mid \kappa_i, \boldsymbol{x}_i, \boldsymbol{\beta}, \lambda)$, $p(\kappa_i \mid y_i, \boldsymbol{x}_i, \boldsymbol{\beta}, \lambda)$ are easy to work with.

Kolmogorov-Gamma random variables

Definition 3. (Kolmogorov-Gamma)

We define $\kappa \sim \mathrm{KG}(b, c)$ as an infinite convolution of independent gammas:

$$\kappa \overset{\mathrm{d}}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{\epsilon_k}{k^2 + c^2/(4\pi^2)}, \quad \epsilon_k \overset{\mathrm{iid}}{\sim} \mathrm{Gamma}(b, 1), \quad k = 1, 2, \ldots$$

- $\pi(\mathrm{KG}(1, 0))^{1/2}$ is same as Kolmogorov distribution [Andrews and Mallows, 1974].
- Comparison with Pólya-Gamma $\omega \sim \mathrm{PG}(b, c)$ [Polson et al., 2013]:

$$\omega \overset{\mathrm{d}}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{\epsilon_k}{(k - 0.5)^2 + c^2/(4\pi^2)}, \quad \epsilon_k \overset{\mathrm{iid}}{\sim} \mathrm{Gamma}(b, 1), \quad k = 1, 2, \ldots$$

Kolmogorov-Gamma augmentation

Theorem 1. (Kolmogorov-Gamma integral identity)

For any $a \in \mathbb{R}$, $b > 0$, and $\eta \in \mathbb{R}$,

$$\frac{(e^\eta)^a}{\{(e^\eta - 1)/\eta\}^b} = e^{(a-b/2)\eta} \int_0^\infty e^{-\kappa\eta^2/2} p_{\mathsf{KG}}(\kappa; b, 0)\mathrm{d}\kappa, \tag{3}$$

where $p_{\mathsf{KG}}(\kappa; b, 0)$ is the density of a $\mathrm{KG}(b, 0)$ random variable.

- Comparision with Pólya-Gamma integral identity for logistic models [Polson et al., 2013]:

$$\frac{(e^\eta)^a}{(e^\eta + 1)^b} = 2^{-b} e^{(a-b/2)\eta} \int_0^\infty e^{-\omega\eta^2/2} p_{\mathsf{PG}}(\omega; b, 0)\mathrm{d}\omega, \tag{4}$$

- Also, $p_{\mathsf{KG}}(\kappa; b, c) \propto \exp(-c^2\kappa/2) p_{\mathsf{KG}}(\kappa; b, 0)$

Conditional conjugacy with latent Gaussian models

- Consider an augmented model with $\eta_i = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}$:

$$p_{\text{aug}}(y_i, \kappa_i \mid \eta_i) = h(y_i, \lambda) \exp\left(\lambda(y_i - 0.5)\eta_i - \kappa_i\eta_i^2/2\right) p_{\mathsf{KG}}(\kappa_i; \lambda, 0), \quad i = 1, \ldots, n$$

- By Theorem 1, it recovers cobin regression model upon marginalizing out κ_i. Also,

$$p(\kappa_i \mid \eta_i, y_i) = p_{\mathsf{KG}}(\kappa_i; \lambda, \eta_i)$$
$$p(y_i \mid \kappa_i, \eta_i) \propto N(\lambda(y_i - 0.5)\kappa_i^{-1}; \eta_i, \kappa_i^{-1}) \text{ in terms of } \eta_i$$

- **Offers conditional conjugacy** for normal prior models & latent Gaussian models
  - (Scale mixture of) Normal prior of $\boldsymbol{\beta}$, e.g. $\eta_i = \boldsymbol{x}(s_i)^{\mathrm{T}}\boldsymbol{\beta}$
  - Spatial regression model, e.g. $\eta_i = \boldsymbol{x}(s_i)^{\mathrm{T}}\boldsymbol{\beta} + u(s_i)$, $u(\cdot) \sim \text{Gaussian Process.}$
- Same strategy can be applied to micobin, replacing λ to λ_i

Blocked Gibbs sampler for cobin regression

- $Y_i \sim \text{cobin}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}, \lambda^{-1})$, $i = 1, \dots, n$ with normal prior $\boldsymbol{\beta} \sim N(\mathbf{0}, \Sigma_\beta)$

---

1. Sample λ from $\text{pr}(\lambda = l \mid \boldsymbol{\beta}) \propto p_\lambda(l) \prod_{i=1}^n p_{\text{cobin}}(y_i; \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}, l^{-1})$, $l = 1, \dots, L$
2. Sample κ_i from $(\kappa_i \mid \lambda, \boldsymbol{\beta}) \overset{\text{ind}}{\sim} \text{KG}(\lambda, \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})$, $i = 1, \dots, n$
3. Sample $\boldsymbol{\beta}$ from $(\boldsymbol{\beta} \mid \lambda, \boldsymbol{\kappa}) \sim N_p(\boldsymbol{m}_\beta, V_\beta)$, where

$$V_\beta^{-1} = X^{\mathrm{T}}\text{diag}(\kappa_1, \dots, \kappa_n)X + \Sigma_\beta^{-1}, \quad \boldsymbol{m}_\beta = V_\beta X^{\mathrm{T}}(y_1\lambda - 0.5\lambda, \dots, y_n\lambda - 0.5\lambda)^{\mathrm{T}}$$

- Some proper prior p_λ for λ and some large upper bound L of λ
- Steps 1,2 jointly updates $(\lambda, \boldsymbol{\kappa})$

Blocked Gibbs sampler for micobin regression

- $Y_i \sim \text{micobin}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}, \psi)$, $i = 1, \ldots, n$ with priors $\boldsymbol{\beta} \sim N(\mathbf{0}, \Sigma_\beta)$ and $\psi \sim \text{Beta}(a_\psi, b_\psi)$

---

1. Sample λ_i from $\text{pr}(\lambda_i = l \mid \boldsymbol{\beta}, \psi) \propto l(1-\psi)^{l-1} p_{\text{cobin}}(y_i; \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}, l^{-1})$, $l = 1, \ldots, L$, $i = 1, \ldots, n$
2. Sample κ_i from $(\kappa_i \mid \lambda_i, \boldsymbol{\beta}) \overset{\text{ind}}{\sim} \text{KG}(\lambda_i, \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})$, $i = 1, \ldots, n$
3. Sample $\boldsymbol{\beta}$ from $(\boldsymbol{\beta} \mid \boldsymbol{\lambda}, \boldsymbol{\kappa}) \sim N_p(\boldsymbol{m}_\beta, V_\beta)$, where
$$V_\beta^{-1} = X^{\mathrm{T}}\text{diag}(\kappa_1, \ldots, \kappa_n)X + \Sigma_\beta^{-1}, \quad \boldsymbol{m}_\beta = V_\beta X^{\mathrm{T}}(y_1\lambda_1 - 0.5\lambda_1, \ldots, y_n\lambda_n - 0.5\lambda_n)^{\mathrm{T}}$$
4. Sample ψ from $(\psi \mid \boldsymbol{\lambda}) \sim \text{Beta}(a_\psi + 2n, b_\psi - n + \sum_{i=1}^n \lambda_i)$

- Steps 1,2 jointly updates $(\lambda, \boldsymbol{\kappa})$, steps 3,4 jointly updates $(\boldsymbol{\beta}, \psi)$

Posterior computation: theory

Theorem 2. (Rapid mixing of Markov chain)

The blocked Gibbs samplers for cobin and micobin regressions are uniformly ergodic. That is, there exist a constant $M > 0$ and $\rho \in [0, 1)$, both independent of initial state, such that $\|P^t(\boldsymbol{\Theta}^{(0)}, \cdot) - \Pi(\cdot)\|_{\mathrm{TV}} \leq M\rho^t$ for all $t \geq 1$.

- Proof by establishing uniform minorization condition, similar to [Choi and Hobert, 2013]
- Guarantees the existence of CLT for Monte Carlo averages of functions of $\boldsymbol{\beta}$
- Strong result for micobin since likelihood is not log-concave

Sampling Kolmogorov-Gamma random variable

$$\kappa \sim \mathrm{KG}(b,c) \iff \kappa \overset{\mathrm{d}}{=} \frac{1}{2\pi^2}\sum_{k=1}^{\infty} \frac{\epsilon_k}{k^2 + c^2/(4\pi^2)}, \quad \epsilon_k \overset{\mathrm{iid}}{\sim} \mathrm{Gamma}(b,1), \quad k = 1, 2, \ldots$$

- Truncating infinite sum of independent gammas: slow, prone to truncation error
- We have $\mathrm{KG}(\lambda, c) \overset{\mathrm{d}}{=} \sum_{l=1}^{\lambda} \mathrm{KG}(1,c)$ for integer λ
- Exact sampling of $\mathrm{KG}(1,c)$: alternating series method [Devroye, 1986]
- Density of $\mathrm{KG}(1,c)$: $p_{\mathsf{KG}}(x;1,c) = \sum_{n=0}^{\infty}(-1)^n a_n(x;c,t)$ with cutoff $t \in (0.0234, 0.25)$,

$$a_n(x;c,t) = \begin{cases} \{\sinh(c/2)/(c/2)\}\exp(-c^2x/2)a_n^L(x), & 0 < x < t, \\ \{\sinh(c/2)/(c/2)\}\exp(-c^2x/2)a_n^R(x), & t \le x \end{cases} \tag{5}$$

- $a_n^L(x)$, $a_n^R(x)$ derived from dual density representation of Kolmogorov r.v. [Feller, 1948]

Sampling Kolmogorov-Gamma $(1, c)$ random variable

1. For $A^L(c,t) = \int_0^t a_0(x;c,t)\mathrm{d}x$ and $A^R(c,t) = \int_t^\infty a_0(x;c,t)\mathrm{d}x$, propose

$$X \sim \begin{cases} \mathrm{GIG}(-1.5, c^2, 1/4)1(0 < X < t) & \text{with prob. } A^L(c,t)/\{A^L(c,t) + A^R(c,t)\} \\ \mathrm{Exp}(c^2/2 + 2\pi^2)1(t \leq X) & \text{with prob. } A^R(c,t)/\{A^L(c,t) + A^R(c,t)\} \end{cases} \tag{6}$$

2. Generate $U \sim \mathrm{Unif}(0, a_0(X;c,t))$
3. Repeat until $U \leq \sum_{n=0}^m (-1)^n a_n(X;c,t)$ (odd m) or $U > \sum_{n=0}^m (-1)^n a_n(X;c,t)$ (even m)
4. Accept X if m is odd, repeat from step 1 again if m is even.

Proposition (KG sampler is fast)

Using the best cutoff point $t^* \approx 0.050239$, the expected number of outer loop & inter loop iterations are bounded above by 1.1456 and 1.1275 for any given c.

Simulation 1: consistency of point estimate

Proposition (consistency of cobin MLE under potential misspecification)

If the mean structure $E(y_i \mid x_i) = g^{-1}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})$ is correctly specified, the solution $\hat{\boldsymbol{\beta}}$ of the cobin regression likelihood equations (2) is **consistent** [Gourieroux et al., 1984]

- Data generation with cobit/logit link functions and 4 different distributions
  - $Y_i \sim \text{beta}(\mu_i, \phi)$, $g(\mu_i) = \beta_0 + \beta_1 x_i$
  - $Y_i \sim \text{cobin}((B')^{-1}(\mu_i), \lambda^{-1})$, $g(\mu_i) = \beta_0 + \beta_1 x_i$
  - $Y_i \sim \text{beta rectangular}(\mu_i, \alpha, \phi) = w_i \text{beta}(\tilde{\mu}_i, \phi) + (1 - w_i)\text{unif}(0, 1)$, $g(\mu_i) = \beta_0 + \beta_1 x_i$
  - $Y_i \sim 0.25\text{beta}(\mu_i - \epsilon_i, \phi) + 0.5\text{beta}(\mu_i, \phi) + 0.25\text{beta}(\mu_i + \epsilon_i, \phi)$, $g(\mu_i) = \beta_0 + \beta_1 x_i$
- Correct link & mean structure $g(E(Y_i \mid x_i)) = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}$, but distribution can be misspecified
- $n \in \{100, 400, 1600\}$, $(\beta_0^{\text{true}}, \beta_1^{\text{true}}) = (0, 1)$, compare betareg $\hat{\boldsymbol{\beta}}$ and cobinreg $\hat{\boldsymbol{\beta}}$

Simulation 1 results

			Beta		Cobin		Beta rectangular		Mixture of beta	
Link	Method	n	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
cobit	beta regression	100	**0.004**	**0.106**	-0.024	0.099	-0.061	0.153	-0.030	0.094
		400	-0.003	**0.054**	-0.030	0.057	-0.069	0.099	-0.037	0.058
		1600	0.002	**0.026**	-0.030	0.038	-0.076	0.084	-0.036	0.042
	cobin regression	100	0.005	0.113	**0.003**	0.099	**0.013**	**0.134**	**0.006**	**0.092**
		400	**-0.002**	0.056	**-0.001**	**0.049**	**0.006**	**0.070**	**-0.001**	**0.046**
		1600	0.002	0.027	**0.000**	**0.023**	**-0.001**	**0.035**	**0.001**	**0.022**
logit	beta regression	100	**0.003**	**0.084**	-0.043	0.080	-0.054	0.117	-0.041	0.074
		400	**0.000**	**0.042**	-0.047	0.059	-0.059	0.080	-0.046	0.055
		1600	0.000	**0.021**	-0.045	0.048	-0.062	0.068	-0.046	0.048
	cobin regression	100	0.015	0.101	**0.005**	**0.066**	**0.020**	**0.116**	**0.005**	**0.067**
		400	0.004	0.051	**0.000**	**0.035**	**0.007**	**0.062**	**0.000**	**0.033**
		1600	0.000	0.026	**0.001**	**0.016**	**0.001**	**0.032**	**0.001**	**0.016**

- Cobin regression $\hat{\beta}$ is consistent even under the misspecified distribution

Simulation 2: spatial regression

- Resembling spatially indexed MMI data $Y(s_i)$
- Data generation: $Y(s_i) \sim \text{beta rectangular}(\mu_i, \alpha, \phi)$, locations $s_i \in [0,1]^2$ uniformly, and

$$g_{\text{cobit}}(\mu_i) = \beta_0 + \beta_1 x(s_i) + u(s_i), \quad u(\cdot) \sim \text{mean zero Gaussian process}$$

- $(\beta_0^{\text{true}}, \beta_1^{\text{true}}) = (0, 1)$, $\rho \in \{0.1, 0.2\}$ (spatial dependence)
- Fit data with (1) spatial beta, (2) spatial cobin, (3) spatial micobin regression models
  - **All models are misspecified**
- `Stan` for spatial beta; Gibbs sampler for spatial cobin/micobin; 5000 MCMC samples.
- Compare (1) Inference of β, (2) predictive performance, (3) sampling performance

Simulation 2 results

ρ	Method	$(n_{\text{train}}, n_{\text{test}})$	Inference ($\hat{\beta}_1$)		Prediction		Sampling ($\boldsymbol{\beta}$)	
			Bias	RMSE	negtestLL	MSPE$\times 10^2$	mESS	time (min)
0.1	beta regression	(200, 50)	-0.048	0.118	-0.325	0.427	919.8	44.5
		(400, 100)	-0.052	0.089	-0.354	0.345	978.7	437.7
	cobin regression	(200, 50)	**0.005**	**0.093**	-0.340	0.388	2791.3	2.0
		(400, 100)	**0.005**	**0.067**	-0.372	0.323	3220.9	11.2
	micobin regression	(200, 50)	0.034	0.099	**-0.367**	**0.373**	1908.4	2.4
		(400, 100)	0.037	0.074	**-0.394**	**0.312**	2137.5	11.7
0.2	beta regression	(200, 50)	-0.065	0.120	-0.320	0.329	1187.2	96.3
		(400, 100)	-0.052	0.095	-0.350	0.248	808.0	933.4
	cobin regression	(200, 50)	**0.000**	**0.088**	-0.346	0.306	3366.0	2.2
		(400, 100)	**0.013**	**0.078**	-0.370	0.233	3663.9	12.1
	micobin regression	(200, 50)	0.039	0.092	**-0.373**	**0.293**	2265.3	2.2
		(400, 100)	0.050	0.091	**-0.395**	**0.226**	2575.4	12.7

Monte Carlo standard errors are all less than 0.015 for negtestLL, 0.013 for MSPE, 127.2 for mESS.

- Cobin gives lowest bias/RMSE of $\hat{\beta}_1$, micobin achieves best predictive performance
- Multivariate ESS per time: cobin and micobin better than 40x or more

Benthic macroinvertebrate multimetric index (MMI)

- Index between 0 and 100, scaled by 0.01
- a.k.a. index of biotic integrity [Karr, 1991]
- Combines various macroinvertebrate assemblages attributes
  - Taxonomic composition, richness, presence of sensitive species (e.g. mayfly)...
- Higher MMI indicates a healthier and diverse benthic macroinvertebrate community
- Data from 2017 NLA survey, 950 lakes
- One lake had a 0 MMI value, set $n = 949$ for comparison with beta regression

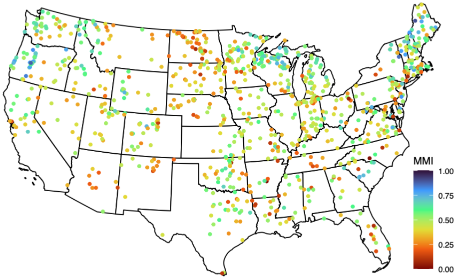


Figure: MMI of 949 US lakes

Lake watershed covariate

- Data source: LakeCat [Hill et al., 2018]
- Various lake catchment/watershed information of 300k+ US lakes
  - For prediction only, we use 55215 lakes with surface area $> 40,000m^2$
- 9 watershed covariates, log transformed:
  - agkffact (soil erodibility), bfi (base flow index), conif (coniferous forest cover)
  - cbnf (cultivated N fixation), crophay (crop&hay land cover), fert (synthetic N fertilizer use), manure (manure application)
  - urbmdhi (medium/high density urban land cover), pestic97 (1997 pesticide use)

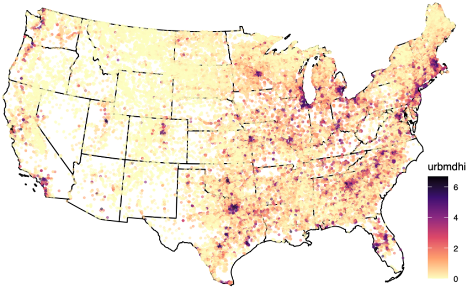


Figure: medium/high density urban land cover percentage, log transformed

Model fit

- Fit three different spatial regression models (beta, cobin, micobin) with

$$g_{\mathrm{cobit}}(\mu_i) = \boldsymbol{x}(s_i)^{\mathrm{T}}\boldsymbol{\beta} + u(s_i), \quad u(\cdot) \sim \text{nearest neighbor GP}$$

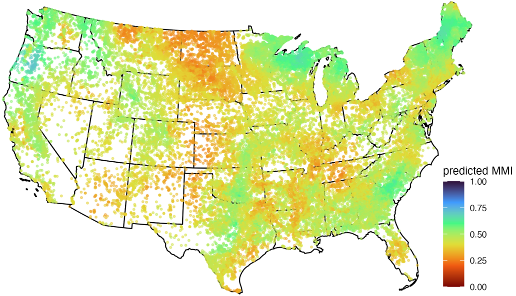
- Employ NNGP [Datta et al., 2016] due to prediction at 50k+ locations
- Prior $\boldsymbol{\beta} \sim N_p(\mathbf{0}, 100^2 I_p)$, half-Cauchy on random effect standard deviation
- `Stan` for spatial beta; Gibbs for spatial cobin/micobin; 5000 MCMC samples, 3 chains
  - Leveraging normal conjugacy via KG augmentation, we jointly update $\boldsymbol{\beta}$ and $\{u(s_i)\}_{i=1}^n$ by partial collapsing [Van Dyk and Park, 2008]
  - Took 2 hr for spatial beta, **5 min for cobin and micobin** per chain
  - mESS/time difference more than 20x.

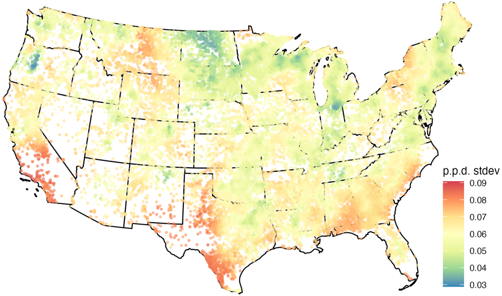
MMI data analysis results: association

	Beta regression		Cobin regression		Micobin regression	
Variable	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Intercept	-2.363	(-4.160, -0.553)	-2.106	(-3.859, -0.345)	-1.797	(-3.551, -0.085)
agkffact	-2.586	(-5.584, 0.330)	**-2.888**	**(-5.714, -0.003)**	**-3.457**	**(-6.113, -0.800)**
bfi	**0.343**	**(0.016, 0.672)**	0.293	(-0.022, 0.614)	0.229	(-0.082, 0.548)
cbnf	0.165	(-0.081, 0.412)	0.182	(-0.055, 0.420)	0.191	(-0.035, 0.425)
conif	0.081	(-0.002, 0.164)	**0.093**	**(0.011, 0.176)**	**0.123**	**(0.044, 0.203)**
crophay	-0.079	(-0.250, 0.091)	-0.063	(-0.231, 0.106)	-0.054	(-0.213, 0.105)
fert	-0.073	(-0.310, 0.158)	-0.092	(-0.323, 0.132)	-0.082	(-0.300, 0.138)
manure	-0.048	(-0.202, 0.102)	-0.036	(-0.182, 0.115)	-0.029	(-0.173, 0.118)
pestic97	-0.014	(-0.106, 0.075)	-0.021	(-0.108, 0.067)	-0.025	(-0.108, 0.059)
urbmdhi	**-0.181**	**(-0.288, -0.076)**	**-0.170**	**(-0.273, -0.067)**	**-0.142**	**(-0.243, -0.041)**

- WAIC: -1093.4 (beta), -1103.5 (cobin), **-1119.3 (micobin)**
- Selected variables based on 95% CI are different for beta

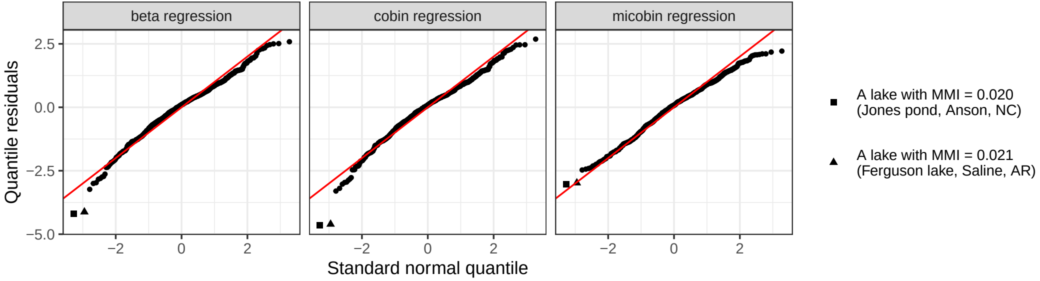
MMI data analysis results: prediction





- Left: predicted MMI, Right: stdev of predicted MMI

MMI data analysis results: goodness of fit



- Quantile residual plot [Dunn and Smyth, 1996]: $\Phi^{-1}(F(y_i \mid \hat{\mu}_i, \hat{\phi}))$ against normal quantiles
- Two influential observations with the lowest MMI values of 0.02 and 0.021
- Re-run the analysis, removing those two lakes

MMI data analysis results: sensitivity analysis

$(n = 947)$	Beta regression		Cobin regression		Micobin regression	
Variable	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
(Intercept)	-1.829	(-3.621, -0.070)	-1.756	(-3.531, 0.006)	-1.741	(-3.520, 0.018)
agkffact	**-3.088**	**(-5.982, -0.206)**	**-3.150**	**(-6.019, -0.258)**	**-3.494**	**(-6.220, -0.822)**
bfi	0.244	(-0.075, 0.568)	0.228	(-0.088, 0.552)	0.219	(-0.097, 0.540)
cbnf	0.191	(-0.044, 0.430)	0.200	(-0.035, 0.437)	0.196	(–0.034, 0.424)
conif	**0.096**	**(0.014, 0.175)**	**0.103**	**(0.021, 0.183)**	**0.125**	**(0.045, 0.204)**
crophay	-0.057	(-0.223, 0.110)	-0.053	(-0.218, 0.114)	-0.050	(-0.210, 0.110)
fert	-0.096	(-0.327, 0.135)	-0.104	(-0.329, 0.122)	-0.089	(-0.316, 0.135)
manure	-0.001	(-0.148, 0.148)	-0.009	(-0.157, 0.138)	-0.022	(-0.167, 0.122)
pestic97	-0.031	(-0.118, 0.057)	-0.030	(-0.119, 0.057)	-0.027	(-0.110, 0.055)
urbmdhi	**-0.180**	**(-0.283, -0.076)**	**-0.169**	**(-0.275, -0.064)**	**-0.143**	**(-0.242, -0.043)**
Change	$\Vert\hat{\boldsymbol{\beta}}^{(949)} - \hat{\boldsymbol{\beta}}^{(947)}\Vert_2 = 0.743$		$\Vert\hat{\boldsymbol{\beta}}^{(949)} - \hat{\boldsymbol{\beta}}^{(947)}\Vert_2 = 0.444$		$\Vert\hat{\boldsymbol{\beta}}^{(949)} - \hat{\boldsymbol{\beta}}^{(947)}\Vert_2 = 0.069$	

- Beta regression results changed, cobin/micobin results remained same
- Recall that beta score function is unbounded in y

MMI data analysis results: sensitivity analysis

	Micobin regression ($n = 950$)		Micobin regression ($n = 949$)	
Variable	Estimate	95% CI	Estimate	95% CI
(Intercept)	-1.758	(-3.517, -0.04)	-1.797	(-3.551, -0.085)
agkffact	**-3.456**	**(-6.175, -0.794)**	**-3.457**	**(-6.113, -0.800)**
bfi	0.219	(-0.100, 0.537)	0.229	(-0.082, 0.548)
cbnf	0.187	(-0.040, 0.415)	0.191	(-0.035, 0.425)
conif	**0.128**	**(0.048, 0.208)**	**0.123**	**(0.044, 0.203)**
crophay	-0.060	(-0.222, 0.101)	-0.054	(-0.213, 0.105)
fert	-0.071	(-0.296, 0.13)	-0.082	(-0.300, 0.138)
manure	-0.031	(-0.178, 0.115)	-0.029	(-0.173, 0.118)
pestic97	-0.023	(-0.106, 0.059)	-0.025	(-0.108, 0.059)
urbmdhi	**-0.141**	**(-0.243, -0.038)**	**-0.142**	**(-0.243, -0.041)**

- Micobin does not require removing a datum with MMI = 0
- Result almost unchanged

- Preprint: `https://arxiv.org/abs/2504.15269`
- Reproducing code: `https://github.com/changwoo-lee/cobin-reproduce`
- R package ”cobin”: `https://github.com/changwoo-lee/cobin`

Thank you!

References I

Andrews, D. F. and Mallows, C. L. (1974).
Scale mixtures of normal distributions.
J. R. Statist. Soc. B, 36(1):99–102.

Bates, G. E. (1955).
Joint distributions of time intervals for the occurrence of successive accidents in a generalized Polya scheme.
Annals of Mathematical Statistics, 26:705–720.

Berggren, N., Daunfeldt, S.-O., and Hellström, J. (2014).
Social trust and central-bank independence.
Eur. J. Polit. Econ., 34:425–439.

Bharti, D. K., Pawar, P. Y., Edgecombe, G. D., and Joshi, J. (2023).
Genetic diversity varies with species traits and latitude in predatory soil arthropods (Myriapoda: Chilopoda).
Glob. Ecol. Biogeogr., 32(9):1508–1521.

Blasco-Moreno, A., Pérez-Casany, M., Puig, P., Morante, M., and Castells, E. (2019).
What does a zero mean? Understanding false, random and structural zeros in ecology.
Methods Ecol. Evol., 10(7):949–959.

References II

Choi, H. M. and Hobert, J. P. (2013).
The Polya-Gamma Gibbs sampler for Bayesian logistic regression is uniformly ergodic.
Electron. J. Stat., 7:2054–2064.

Cribari-Neto, F. and Zeileis, A. (2010).
Beta Regression in R.
J. Stat. Softw., 34(2):1–24.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016).
Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets.
J. Am. Stat. Assoc., 111(514):800–812.

de Vargas Ribeiro, F., Pessarrodona, A., Tucket, C., Mulders, Y., Pereira, R. C., and Wernberg, T. (2022).
Shield wall: Kelps are the last stand against corals in tropicalized reefs.
Funct. Ecol., 36(10):2445–2455.

Devroye, L. (1986).
Non-Uniform Random Variate Generation.
Springer New York.

References III

Dunn, P. K. and Smyth, G. K. (1996).
Randomized quantile residuals.
J. Comput. Graph. Stat., 5(3):236.

Feller, W. (1948).
On the Kolmogorov-Smirnov limit theorems for empirical distributions.
Ann. Math. Stat., 19(2):177–189.

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

Gourieroux, C., Monfort, A., and Trognon, A. (1984).
Pseudo maximum likelihood methods: Theory.
Econometrica, 52(3):681.

Hill, R. A., Weber, M. H., Debbout, R. M., Leibowitz, S. G., and Olsen, A. R. (2018).
The Lake-Catchment (LakeCat) Dataset: characterizing landscape features for lake basins within the conterminous USA.
Freshw. Sci., 37:208–221.

References IV

Jørgensen, B. (1987).
Exponential dispersion models.
J. R. Statist. Soc. B, 49(2):127–145.

Karr, J. R. (1991).
Biological integrity: A long-neglected aspect of water resource management.
Ecol. Appl., 1(1):66–84.

Korhonen, P., Hui, F. K. C., Niku, J., Taskinen, S., and van der Veen, B. (2024).
A comparison of joint species distribution models for percent cover data.
Methods Ecol. Evol., 15(12):2359–2372.

Kosmidis, I. and Zeileis, A. (2024).
Extended-support beta regression for $[0, 1]$ responses.
arXiv preprint arXiv:2409.07233.

Kubinec, R. (2023).
Ordered beta regression: A parsimonious, well-fitting model for continuous data with lower and upper bounds.
Polit. Anal., 31(4):519–536.

References V

Lindholm, M., Alahuhta, J., Heino, J., and Toivonen, H. (2021).
Temporal beta diversity of lake plants is determined by concomitant changes in environmental factors across decades.
J. Ecol., 109(2):819–832.

Loaiza-Ganem, G. and Cunningham, J. (2019).
The continuous Bernoulli: fixing a pervasive error in variational autoencoders.
Adv. Neural Inf. Process. Syst., 32:13287–13297.

Nelder, J. A. and Wedderburn, R. W. M. (1972).
Generalized linear models.
J. R. Statist. Soc. A, 135(3):370.

Peplonska, B., Bukowska, A., Sobala, W., Reszka, E., Gromadzinska, J., Wasowicz, W., Lie, J. A., Kjuus, H., and Ursin, G. (2012).
Rotating night shift work and mammographic density.
Cancer Epidemiol. Biomarkers Prev., 21(7):1028–1037.

Polson, N. G., Scott, J. G., and Windle, J. (2013).
Bayesian inference for logistic models using Pólya–Gamma latent variables.
J. Am. Statist. Assoc., 108(504):1339–1349.

References VI

Qiao, J., Chu, L., Li, Y., Chu, T., Xie, N., and Yan, Y. (2025).
Unraveling spatial patterns and drivers of fish ecological uniqueness in subtropical streams.
Ecol. Evol., 15(4):e71112.

Rolls, R. J., Wolfenden, B., Heino, J., Butler, G. L., and Thiem, J. D. (2023).
Scale dependency in fish beta diversity–hydrology linkages in lowland rivers.
J. Biogeogr., 50(10):1692–1709.

Smithson, M. and Verkuilen, J. (2006).
A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables.
Psychol. Methods, 11(1):54–71.

U.S. Environmental Protection Agency (2022).
National lakes assessment 2017: Technical support document. EPA 841-R-22-001.
https://www.epa.gov/national-aquatic-resource-surveys/nla.

Van Dyk, D. A. and Park, T. (2008).
Partially collapsed Gibbs samplers: Theory and methods.
J. Am. Statist. Assoc., 103(482):790–796.

References VII

van Strien, A. J., Irvine, K. M., and Retel, C. (2024).
Trends in plant cover derived from vegetation plot data using ordinal zero-augmented beta regression.
J. Veg. Sci., 35(4).

Wang, C. and Blei, D. M. (2018).
A general method for robust Bayesian modeling.
Bayesian Anal., 13(4):1163–1191.

Warton, D. I. and Hui, F. K. C. (2011).
The arcsine is asinine: the analysis of proportions in ecology.
Ecology, 92(1):3–10.