

# Appendix S1

## Generalized joint attribute modeling for biodiversity analysis: Median-zero, multivariate, multifarious data

JAMES S. CLARK<sup>1</sup>, DIANA NEMERGUT<sup>4</sup>, BIJAN SEYEDNASROLLAH<sup>1</sup>, PHILLIP TURNER<sup>2</sup>,  
and STACY ZHANG<sup>2</sup>

October 29, 2016

<sup>1</sup> *Nicholas School of the Environment, Duke University, Durham NC 27708, USA*

<sup>2</sup> *Division of Marine Science and Conservation, Nicholas School of the Environment, Duke University, Beaufort, NC 28516, USA*

<sup>3</sup> *Department of Statistical Science, Duke University, Durham, NC 27708-0251*

<sup>4</sup> *Department of Biology, Duke University, Durham, NC 27708*

## Contents

<b>1</b>	<b>Summary</b>	<b>1</b>
<b>2</b>	<b>Additional model background</b>	<b>1</b>
2.1	Graphical interpretation . . . . .	1
2.2	Likelihood for censored values and covariance . . . . .	1
2.3	Prior distributions . . . . .	3
2.4	Interval censoring by data type . . . . .	3
2.5	Composition data . . . . .	4
2.6	Missclassification, including zero inflation . . . . .	5
<b>3</b>	<b>Computation notes</b>	<b>7</b>
<b>4</b>	<b>Prediction notes</b>	<b>7</b>
4.1	Distribution theory . . . . .	7
4.2	Inverse prediction . . . . .	8
4.3	Missing $X$ . . . . .	8
4.4	Missing $Y$ . . . . .	9
<b>5</b>	<b>Applications</b>	<b>9</b>
5.1	FIA . . . . .	9
5.2	EMP . . . . .	10
<b>6</b>	<b>Additional references</b>	<b>11</b>

# 1 Summary

Here we provide background on the model, computation, and data. Citations not included in the main text are given in Section 6. Additional information is available as vignettes for the R package `gjam`: <https://cran.rstudio.com/web/packages/gjam/index.html>.

## 2 Additional model background

### 2.1 Graphical interpretation

The interpretation of discrete observations is extended here in graphical form. The regression (dashed line) in Figure S1a, c is the model for latent  $W|x$ . The model 'generates' a continuous  $W$ , then a discrete observation  $Y = 3$ , because the realized  $W$  lies within the interval  $(p_3, p_4]$  (Fig. S1b). Model fitting inverts these steps, starting with the observation  $Y = 3$  (Fig. S1d), which is assigned by the partition to an interval for  $W$  (Fig. S1c). Note that values of  $W < 0$  are assigned by  $(p_0, p_1] = (-\infty, 0]$  to  $Y = 0$ .

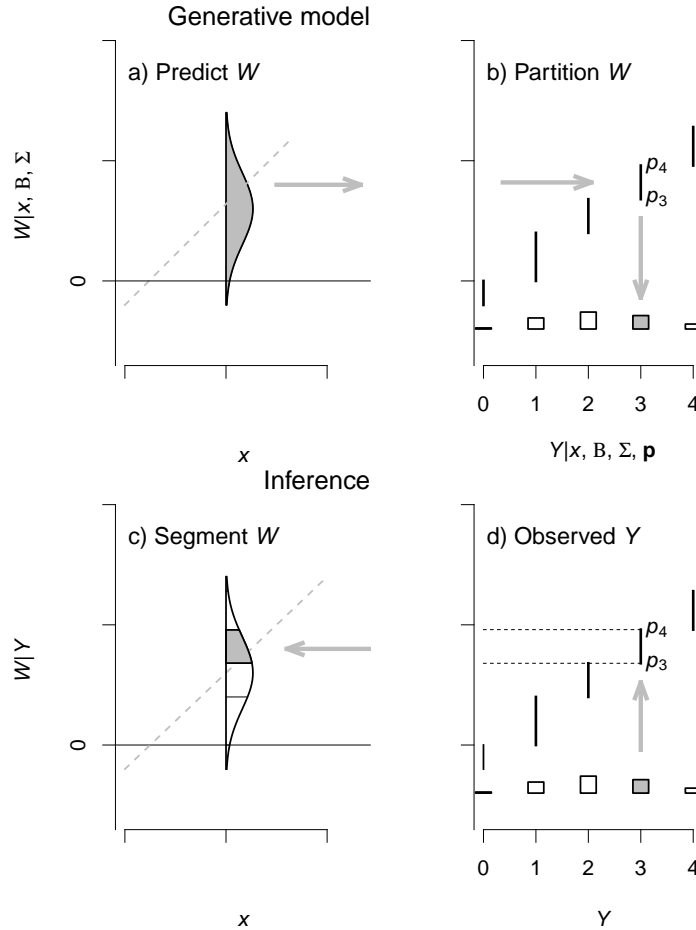


Figure S1: Graphical relationships between continuous  $W$  and discrete  $Y$ , highlighted for an observed value of  $Y = 3$ . The regression is shown as a dashed line in (a) and (c). For a given value of  $x$  there is a distribution of values for  $W$ . The partition, shown as vertical lines in (b) and (d), segments  $W$  into discrete values of  $Y$ . As a generative model (a, b),  $W$  is random given  $x, \mathbf{B}, \Sigma$ . For inference (c, d), a distribution of observed  $Y$  (d) maps to the truncated distribution for  $W$  (c).

## 2.2 Likelihood for censored values and covariance

Here we use the likelihood function to relate the variance in an observation to the width of a partition interval. To simplify notation we assume  $\forall(i, s) : E_{is} = E = 1$ , and we omit subscript  $i$  on  $w$  and  $y$ . For discrete, and thus, interval-censored, observations the joint probability of two species  $(s, s')$  within intervals  $(k, k')$  is

$$[y_s = k, y_{s'} = k'] = (2\pi)^{-S/2} |\Sigma|^{-1/2} \int_{p_{s',k'}}^{p_{s',k'+1}} \int_{p_{s,k}}^{p_{s,k+1}} \exp\left[-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{w}-\boldsymbol{\mu})\right] dw_s dw_{s'} \quad (\text{S1})$$

Obviously, the wider the interval, the greater the probability that  $(w_{is}, w_{is'})$  will lie within it.

There is a contribution to the variance that comes from integrating over the interval. For species  $s$  and interval  $k$  having interval width  $p_{s,k+1} - p_{s,k}$  we have the conditional second moment

$$E[w_s^2 | w_{s'}] = \int_{p_{s,k}}^{p_{s,k+1}} w_s^2 \phi(w_s; \boldsymbol{\mu}_s, \sigma_s) dw_s$$

where  $\phi$  is normal density function with conditional mean and standard deviation  $\boldsymbol{\mu}_s, \sigma_s$ . For a narrow interval this integral is proportional to the width of interval  $k$ . The example in Figure 3a of the main text shows this effect on  $y_{is}$ , which influences the weight of observations, while still allowing us to estimate  $\Sigma$ .

## 2.3 Prior distributions

Prior distributions for examples in the main text are non-informative. This is particularly helpful for the covariance, lending stability to Gibbs sampling and avoiding dominance by a prior (Section 3). From Eq. 2 of the main text we include prior distributions,

$$\prod_{i=1}^n MVN(\mathbf{w}_i | \mathbf{B}'\mathbf{x}_i, \Sigma) \times MVN(\text{vec}(\mathbf{B}) | 0, \Sigma \otimes \mathbf{C}) \times |\Sigma|^{-(S+2)/2} \quad (\text{S2})$$

where the prior matrix  $\mathbf{C}$  has infinite variance. The third factor is the Jeffreys prior for the covariance matrix (Yang and Berger 1994, Sun and Berger 2006). Gibbs sampling is discussed in Section 3.

The R package `gjam` (<https://cran.rstudio.com/web/packages/gjam/index.html>) admits truncated prior distributions on  $\mathbf{B}$ , as can be useful when the sign of an effect is known (Clark et al. 2014).

## 2.4 Interval censoring by data type

In this section we provide background on specific data types that involve interval censoring and shown in Table 2 of the main text. We show how the partition links to the indicator function in Eq. 3 of the main text,

$$I_{is,k} = I(p_{is,k} < w_{is} \leq p_{is,k+1}) \quad (\text{S3})$$

We do not repeat here discussion of data types that use this indicator precisely and are described in the main text.

Chib and Greenberg's (1998) probit model is a special case of GJAM for *presence-absence* data, where censoring could be represented as  $\mathcal{I}_{is} = I(w_{is} > 0)^{y_{is}} I(w_{is} \leq 0)^{1-y_{is}}$ .

Zhang et al.'s (2008) model for multivariate *categorical data* fit within the same framework. Let  $w_{is,k}$  refer to the latent value for the  $k^{th}$  level of response  $s$ . For multiple responses,

$$y_{is,k} = \begin{cases} 1, & w_{is,k} > 0 \wedge w_{is,k} = \max_{k'}(w_{is,k'}) \\ 0, & w_{is,k} < \max_{k'}(0, w_{is,k'}) \end{cases} \quad (S4)$$

where  $k' = 1, \dots, K_s - 1$ , and  $K_s$  is the number of levels for response  $s$ . The first line of Eq. S4 specifies that the observed class must not only be positive, but also the largest  $w_{is,k}$  value in the  $K_s - 1$  columns. If all values are negative, then the reference class is the observed class. The matrix  $\mathbf{W}$  has columns for each non-reference level of each categorical response,  $\mathbf{w}_i = (w_{i1,1}, \dots, w_{i1,K_1-1}, \dots, w_{iS,1}, \dots, w_{iS,K_S-1})$ . The  $\sum_s K_s - S$  rows and columns of correlation matrix  $\Sigma$  hold correlations for all combinations of  $s$  and  $k$ . The indicator for correct interval membership could be written for this model as

$$\mathcal{I}_{is} = \prod_{k=1}^{K-1} [I(w_{is,k} > 0) I(w_{is,k} = \max_{k'}(w_{is,k'}))^{y_{is,k}} \times I(w_{is,k} < \max_{k'}(0, w_{is,k'}))^{1-y_{is,k}}$$

For multivariate *ordinal data* Lawrence et al.'s (2008) model is a special case of GJAM. The multivariate partition has elements  $p_{s,k}$  that differ for each species, both in number and location. Because there is no scale, all but the first two and the last elements must be inferred. The indicator function (S3) applies to ordinal counts.

The partitions for other data types follow Eq. (S3) and are discussed in the main text.

## 2.5 Composition data

Here we discuss the link between  $W$  and  $Y$  for composition data. The latent  $W$  must be supported over  $(-\infty, \infty)$ . A generative model can predict  $W$  anywhere in this range, which is partitioned as  $\mathbf{p}_{is} = (-\infty, 0, 1, \infty)$ . Recall that fractional composition data have the sum-to-one constraint,  $\sum_s^S y_{is} = 1$  and support for  $y_{is}$  over  $[0, 1]$ , with point masses for absent species and when an observation is dominated by a single species. Count composition data are integer values, but because they have interpretation only in a relative sense, they too are modeled on the  $[0, 1]$  scale. In both cases there is a redundant column in  $\mathbf{Y}$ , defined here to be the reference species  $S$ . To discuss this common issue for both types of composition data we introduce the variable  $\eta_{is} \in [0, 1]$  as the relative scale that applies in either case,  $\sum_s^S \eta_{is} = 1$ . Our goal is to link the latent scale  $w_{is} \in (-\infty, \infty)$  to  $\eta \in [0, 1]$ , which can have point masses at zero and one, but cannot be negative or greater than one. We make this connection through the reference species, arbitrarily the last species  $S$ .

For prediction, the sum of non-reference values must be less than or equal to one. Here is the sum and inequality for a latent vector  $\mathbf{w}_i$ :

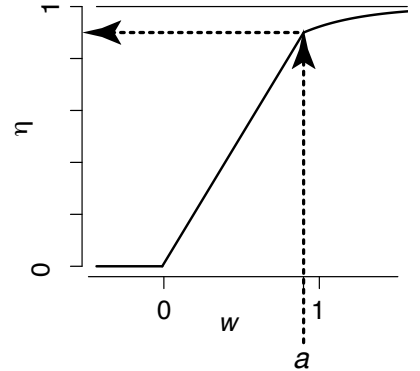


Figure S2: The nearly-linear link function  $g^{-1}$  for composition data. Up to a constant  $a$  the link is linear.

$$\tilde{w}_i = \sum_{s \neq S} [I(0 < w_{is} < 1)w_{is} + I(w_{is} > 1)] \leq 1 \quad (\text{S5})$$

Because the  $w_{is}$  are constrained by data, and data determine parameter values,  $\mathbf{w}_i$  predicted by Eq. 2 of the main text will typically satisfy this inequality, but not always, and not in theory. For cases where predicted  $\mathbf{w}_i$  do not satisfy Eq. S5, we require a link function, one that maps a new vector  $\boldsymbol{\eta}_i \rightarrow \mathbf{w}_i$  with the sum-to-one constraint on  $\boldsymbol{\eta}$ . Additionally, we want a nearly linear relationship between  $\eta_{is}$  and  $w_{is}$ , again, with the goal of modeling data on the scale where they would be interpreted, in this case  $[0, 1]$ . Here we describe a link function that meets these criteria.

The sum of the non-reference species for the composition variable is

$$\tilde{\eta}_i = \sum_{s \neq S} \eta_{is} \leq 1 \quad (\text{S6})$$

Near-identity is obtained if  $\eta_{is} = w_{is}$  in the range  $0 < \tilde{w}_i < a$  for a value of  $a$  that is close to 1. In other words, the two are equivalent whenever the sum of non-reference species is not too large, specifically,  $\tilde{\eta}_i < a$  (Fig. S2). When  $\tilde{w}_i > a$ , the scales are not precisely equivalent,  $\tilde{w}_i > \tilde{\eta}_i$ , and the  $\mathbf{w}_i$  vector must contract. The inverse link function from  $\mathbf{w}_i \rightarrow \boldsymbol{\eta}_i$  is

$$g^{-1}(w_{is}; a) = \begin{cases} 0, & w_{is} \leq 0 \\ w_{is}, & \tilde{w}_i < a, w_{is} > 0 \\ D_i w_{is}, & \tilde{w}_i \geq a, w_{is} > 0 \end{cases}$$

where the compression factor is  $D_i = \tilde{w}_i^{-1}[1 - (1 - a)^{\tilde{w}_i/a}]$ . The link function ( $\boldsymbol{\eta}_i \rightarrow \mathbf{w}_i$ ) balances with a proportionate expansion

$$g(\eta_{is}; a) = \begin{cases} < 0, & \tilde{\eta}_i = 0 \\ \eta_{is}, & \tilde{\eta}_i < a \\ C_i \eta_{is}, & a < \tilde{\eta}_i \leq 1 \end{cases}$$

the expansion factor being  $C_i = \tilde{\eta}_i^{-1} a^{\frac{\log(1-\tilde{\eta})}{\log(1-a)}}$ .

Figure S2 shows how the nonlinear portion of the link function applies where the sum of non-reference species exceeds  $a$  for count composition data. When this occurs the link is not a precise identity. It is close, because data  $\mathbf{y}_i$  tend to bound  $\tilde{w}_i$  away from 1, and  $a$  can be specified close to 1. When  $\tilde{w}_i > a$ , the degree of compression is still small.

As with other data types in GJAM, this model for composition data is generative—we can predict  $\mathbf{w}_i$  from  $\mathbf{B}'\mathbf{x}_i$ , then link  $\mathbf{w}_i \rightarrow \boldsymbol{\eta}_i$ . For fractional composition data  $\mathbf{y}_i = \boldsymbol{\eta}_i$ . For count composition data, predicted  $\boldsymbol{\eta}_i$  translates to  $\mathbf{y}_i$  through the partition (Fig. 3f of main text). Extensive simulation shows that GJAM predicts composition data (e.g., Figure 7 of the main text).

## 2.6 Missclassification, including zero inflation

If the discrete state can be miss-classified the model allows for uncertain  $\mathbf{z}_i$ . Then  $\mathbf{z}_i$  in Eq. 2 of the main text moves to the left of the bar,

$$\begin{aligned} \mathbf{w}_i, \mathbf{z}_i | \mathbf{x}_i, \mathbf{y}_i, \mathbf{E}_i &\sim MVN(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}) \times \prod_{s=1}^S \mathcal{I}_{is} \\ \mathcal{I}_{is} &= \prod_{k \in \mathcal{C}} I_{is,k}^{I(z_{is}=k)} (1 - I_{is,k})^{I(z_{is} \neq k)} \times \psi_k(y_{is}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}) \end{aligned} \quad (\text{S7})$$

The function

$$\psi_k(y_{is}; \mu_{is}, \sigma_{is}) = [z_{is} = k | y_{is}; \mu_{is}, \sigma_{is}] \quad (\text{S8})$$

now appears on the right, being the probability that the true interval is  $z_{is} = k$  given that the observation was assigned interval  $y_{is}$ . Parameters are the conditional mean and variance for response  $s$  in observation  $i$ .  $\psi_k(y_{is})$  depends on the miss-classification error probability (by Bayes Theorem). Miss-classification is the inverse of  $\psi_k(y_{is})$ , i.e., it is the probability that the observation is assigned to an interval  $y_{is}$  that might not be the true interval  $k$ . Again, the most common miss-classification is zero-inflation, described by  $\psi_k(0)$  for true  $k > 0$ .

With zero-inflation not only  $W$ , but also  $Z$  is stochastic. Zero inflation is a specific type of misclassification or observation error, assignment of an observation to a discrete class  $y_{is}$  when the true state  $z_{is}$  belongs to a different class. Misclassification requires a probability of  $z_{is}$  in terms of  $w_{is}$ . From Eq. 2 of the main text  $z_{is}$  and  $w_{is}$  are related as

$$F_{is,k} = [z_{is} = k | \mu_{is}, \sigma_{is}] = \int_{p_{s,k}}^{p_{s,k+1}} \phi(w_{is}; \mu_{is}, \sigma_{is}) dw_{is} \quad (\text{S9})$$

where  $\phi()$  is the normal density function for  $s$  having conditional mean and variance depending on other species. Misclassification error has a probability that an observation is assigned to class  $k'$  when it belongs in class  $k$ ,

$$\theta_{is}(k', k) = [y_{is} = k' | z_{is} = k] \quad (\text{S10})$$

By Bayes Theorem, the probability that the true interval is  $k$  is

$$\psi_{is,k}(k') = [z_{is} = k | y_{is} = k'] = \frac{\theta_{is}(k', k) F_{is,k}}{\sum_k \theta_{is}(k', k) F_{is,k}} \quad (\text{S11})$$

Zero inflation is an example. Consider a continuous response with zeros (Fig. 3b of main text). We wish to inflate the zero class by allowing that it could represent not only the model for  $Z = 0$  but also for  $Y = 0$  when there is detection/sampling error. In the case of two classes, we could have a positive (continuous) class and the zero class associated with  $Y = 0$ ,  $Z = 0$ , and  $W < 0$  (Fig. 3c). A zero observation can obtain because the species cannot survive at a location  $i$ , i.e., for reasons that might be explained by the model  $\mu_{is}$  or because it is missed due to sampling, with probability  $\theta_{is}(0, 1)$  (it is present and, thus, belongs in class 1, but it is not observed and erroneously assigned class 0). For a continuous response with zero inflation the probability that the species cannot survive at the site given observed zero is

$$\psi_{is,0}(0) = [z_{is} = 0 | y_{is} = 0] = \frac{F_{is,0}}{F_{is,0} + (1 - F_{is,0})\theta_{is}(0, 1)} \quad (\text{S12})$$

where

$$F_{is,0} = [w_{is} < 0] = \Phi(\mu_{i,s}, \sigma_s)$$

(Eq. S9). An observed zero has some probability of being non-zero and the probability that the species is simply missed in sampling,

$$[y_{is} = 0] = F_{is,0} + (1 - F_{is,0})\theta_{is}(0, 1)$$

The second term 'inflates' the zeros contributed by the first term. Sampling a latent true class for an observation  $[z_{is}|y_{is} = 1]$  can shift the class to which latent  $w_{is}$  belongs (Fig. 3c) and thus the partition used to sample  $w_{is}$  (Eq. 2).

A number of models for detection error are possible. A beta prior density

$$\theta(0, 1) \sim \text{beta}(q_1, q_2)$$

can have parameter values selected by moment matching. A sample-based prior might reflect levels of effort  $E_i$  associated with samples, detection error declining with effort, such that the mean of  $\theta(0, 1)$  is proportional to  $\exp(-E_i)$ . Effort could be related to plot size, search time, or search area. Selection of prior parameter values  $(q_{1i}, q_{2i})$  to give this prior mean with specified weight lead to the conditional posterior

$$\theta(0, 1) \sim \text{beta}(q_{1i} + \sum_s I(y_{is} = 0)I(z_{is} = 1), q_{2i} + \sum_s I(y_{is} > 0))$$

Alternatively, a species-based prior might reflect crypsis, behavior, or other factors influencing detection probability, such that the mean of  $\theta(0, 1)$  is proportional to  $\exp(-E_s)$ .

A third option is a model-based zero class (Clark et al. 2014). Detection error can have a probit specification,  $\theta_{i,s}(0, 1) = \Phi(\gamma_s \mathbf{u}_i')$ , where  $\mathbf{u}_i$  is a design vector,  $\gamma_s$  is the vector of coefficients, and  $\Phi()$  is the standard normal distribution function.

A fourth option, where  $\mathbf{u}_i$  is taken to be identical to  $\mathbf{x}_i$  (presence and abundance explained by the same predictors), makes parameters proportional to those used to predict abundance,  $\theta_{i,s}(0, 1) = \Phi(\beta_s \mathbf{x}_i' \tau_s)$  (Lambert 1992).

### 3 Computation notes

All latent variables and parameters are drawn directly from conditional posteriors,  $\Sigma|(\mathbf{W}, \mathbf{B})$ ,  $\mathbf{B}|(\Sigma, \mathbf{W})$ , and, for unknown partition (ordinal variables),  $\mathcal{P}|\mathbf{Z}, \mathbf{W}$ . Here are conditional distributions:

$$\begin{aligned} \text{vec}(\mathbf{B}) &\sim \text{MVN}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}, \Sigma \otimes (\mathbf{X}'\mathbf{X})^{-1}) \\ \Sigma &\sim IW(n - Q + S - 1, \mathbf{W}'\mathbf{D}\mathbf{W}) \end{aligned}$$

where  $\mathbf{D} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ . Here is the conditional distribution for the ordinal partition:

$$p_{sk} \sim \text{unif}(\max_i(w_{is}|z_{is} = k), \min_i(w_{is}|z_{is} = k + 1))$$

(Lawrence et al. 2008). Latent variables for ordinal, presence-absence, and categorical data are drawn on the correlation scale,  $\mathbf{W}|\mathbf{R}, \boldsymbol{\alpha}, \mathbf{P}$ , where  $\mathbf{R} = \mathbf{V}^{-1/2}\Sigma\mathbf{V}^{-1/2}$ ,  $\boldsymbol{\alpha} = \mathbf{V}^{-1/2}\boldsymbol{\beta}$ ,  $\mathbf{P} = \mathbf{V}^{-1/2}\mathcal{P}$ , and  $\mathbf{V} = \text{diag}(\Sigma)$ . For other data types, latent variables are drawn on the covariance scale,  $\mathbf{W}|\Sigma, \boldsymbol{\beta}, \mathcal{P}$ . Due to direct sampling, convergence is fast.

## 4 Prediction notes

The prediction methods discussed in the main text follow those of Clark et al. (2011, 2013), with some background for implementation in GJAM summarized here.

### 4.1 Distribution theory

We use two predictive distributions. The environmental component of the response ( $\mathbf{E}$  of Eq. 7 in the main text) makes use of the predictive distribution for response  $\tilde{Y}$ ,

$$[\tilde{Y}|\tilde{X}, X, Y] = \int [\tilde{Y}|\tilde{X}, \hat{\theta}] [\hat{\theta}|X, Y] d\hat{\theta} \quad (\text{S13})$$

The integrand contains the likelihood (Eq. S1) and the posterior distribution for parameters,  $\hat{\theta} = \{\hat{\mathbf{B}}, \hat{\Sigma}, \hat{\mathcal{P}}\}$ . Input  $\tilde{X}$  can equal  $X$  (in-sample prediction) or not (out-of-sample prediction). If we place a prior distribution on  $\tilde{X}$  we can marginalize it as well,

$$[\tilde{Y}|X, Y] = \int [\tilde{Y}|\tilde{X}, X, Y] [\tilde{X}] d\tilde{X} \quad (\text{S14})$$

Whereas Eq. S13 evaluates the prediction for a specific  $\tilde{X}$ , Eq. S14 incorporates variation in  $\tilde{X}$ .

The second predictive distribution is needed for the sensitivity matrix  $\mathbf{F}$  of Eq. 8 of the main text. The inverse prediction for input  $\tilde{X}$  is

$$[\tilde{X}|\tilde{Y}, X, Y] = \frac{[\tilde{Y}|\tilde{X}, X, Y] [\tilde{X}]}{[\tilde{Y}|X, Y]} \quad (\text{S15})$$

We term this the 'inverse prediction' (Clark et al. 2011, 2013).

### 4.2 Inverse prediction

Inverse predictions of  $X$  for linear terms are sampled directly,

$$\tilde{\mathbf{x}}_i | (\mathbf{B}, \Sigma, \mathbf{w}_i) \sim MVN(\mathbf{V}\mathbf{v}, \mathbf{V}) \quad (\text{S16})$$

where

$$\begin{aligned} \mathbf{V}^{-1} &= \mathbf{B}'\Sigma^{-1}\mathbf{B} + \mathbf{U}^{-1} \\ \mathbf{v} &= \mathbf{B}'\Sigma^{-1}\mathbf{w}_i + \mathbf{U}^{-1}\mathbf{u} \end{aligned}$$

and  $\mathbf{U}, \mathbf{u}$  are the prior covariance and mean vector for  $\tilde{\mathbf{x}}_i$ . Non-linear terms (those involved in interactions and squared terms) are sampled using Metropolis.

Multilevel factors become categorical variables upon inversion. The predicted category is the largest predicted value from Eq. S16 for factor levels. Predictions are marginalized over the posterior distribution of parameters and latent states.



### 4.3 Missing $X$

Missing values in  $\mathbf{x}_i$  require conditional prediction, i.e., only for those values that are missing. The missing values  $\mathbf{x}_{i,A}$  are a subset of the vector,  $A \subseteq \{2, \dots, Q\}$  (i.e., not the intercept). The non-missing values  $A^C$  are the complement of  $A$  in observation  $i$  and always includes the intercept. For linear terms, missing values are drawn from

$$\tilde{\mathbf{x}}_{i,A} | (\mathbf{x}_{i,A^C}, \mathbf{B}, \boldsymbol{\Sigma}, \mathbf{w}_i) \sim MVN(\mathbf{V}\mathbf{v}, \mathbf{V}) \quad (\text{S17})$$

where

$$\begin{aligned} \mathbf{V}^{-1} &= \mathbf{B}'_A \boldsymbol{\Sigma}^{-1} \mathbf{B}_A + \mathbf{U}_A^{-1} \\ \mathbf{v} &= \mathbf{B}'_A \boldsymbol{\Sigma}^{-1} (\mathbf{w}_i - \mathbf{B}_{A^C} \mathbf{x}_{i,A^C}) + \mathbf{U}_A^{-1} \mathbf{u}_A \end{aligned}$$

and  $\mathbf{U}_A, \mathbf{u}_A$  are the prior covariance and mean vector for  $\mathbf{x}_{i,A}$ , and  $\mathbf{B}_A$  and  $\mathbf{B}_{A^C}$  are the rows in  $\mathbf{B}$  corresponding to  $A$  and  $A^C$ , respectively. Non-linear missing values (those involved in interactions and squared terms) are sampled using Metropolis.

When one category of a multilevel factor is missing, all are missing,  $A$  is the set of all categories for the factor, and the predicted category is the largest predicted value in the  $\tilde{\mathbf{x}}_{i,A}$  vector from Eq. S17.

Estimates are marginalized over the posterior distribution of parameters and latent states.

### 4.4 Missing $Y$

Missing values in  $\mathbf{Y}$  are predicted conditionally, first  $W|X$ , then translated to  $Y$  by partition  $\mathcal{P}$ . Missing latent variables are drawn conditionally,

$$\tilde{\mathbf{w}}_{i,A} | (\mathbf{w}_{i,(A^C)}, \mathbf{B}, \boldsymbol{\Sigma}, \mathbf{x}_i) \sim MVN(\boldsymbol{\mu}_{is|s^C}, \boldsymbol{\Sigma}_{s|(s^C)})$$

where  $\mathbf{w}_{i,A}, \mathbf{w}_{i,A^C}$  are missing and non-missing responses in  $i$ , and  $\boldsymbol{\mu}_{is|s^C}, \boldsymbol{\Sigma}_{s|(s^C)}$  are the mean vector and covariance for missing responses conditional on non-missing responses. For continuous observations  $Y = W$ . For discrete observations  $y_{is} | (\mathcal{P}, w_{is}) = \sum_k k I(p_k < w_{is} < p_{k+1})$ .

Estimates are marginalized over the posterior distribution of parameters and latent states.

## 5 Applications

### 5.1 FIA

FIA data are publicly available at <http://www.fia.fs.fed.us/tools-data/index.php>. The predictors for the FIA data include climate variables, soils, and local drainage, discussed here:

**Stand age:** Recorded as part of the FIA census for each plot.

**Temperature:** The mean annual temperature (basal area example) and winter temperature (trait example) are from the MODIS land surface temperature. Winter temperature is for the months December, January, and February (Fig. S3c).

**Deficit:** A regional indicator of drought. The hydrothermal deficit is the degree hours accumulated for months with a negative water balance, reflecting the major limitations on net primary productivity, solar radiation, temperature, and moisture. Water balance variables are

temperature  $T_{im}$ , precipitation  $P_{im}$ , and  $PET_{im}$  for location  $i$  and month  $m$ . Months with a negative water balance,

$$D_{im} = \begin{cases} 1, & P_{im} < PET_{im} \\ 0, & P_{im} \geq PET_{im} \end{cases} \quad (\text{S18})$$

scale the degree-hours  $DH_{im} = T_{im} \times DF_{im}$ , where the second factor is the average daylight fraction for month  $m$ . The deficit is

$$D_i = \sum_{m=1}^{12} DH_{im} \times D_{im} \quad (\text{S19})$$

(Fig. S3b).

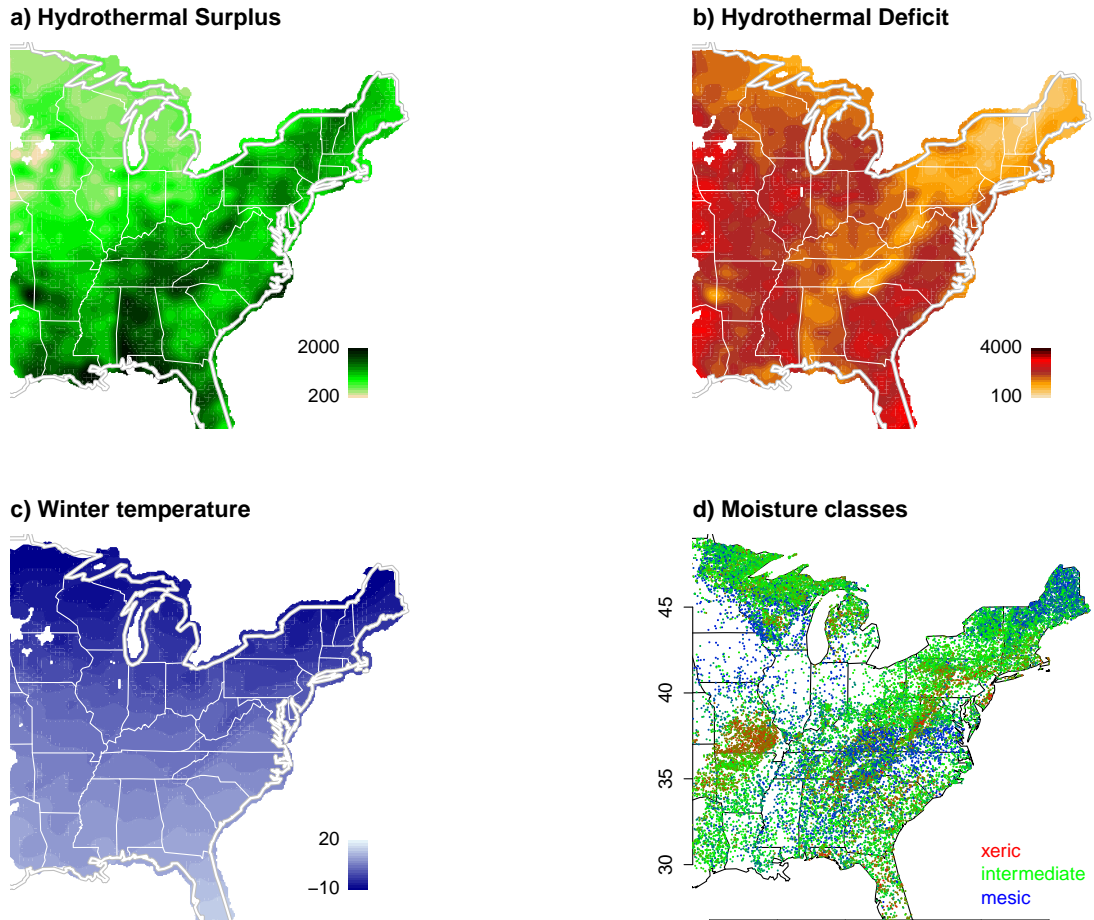


Figure S3: Climate and moisture variables used as predictors.

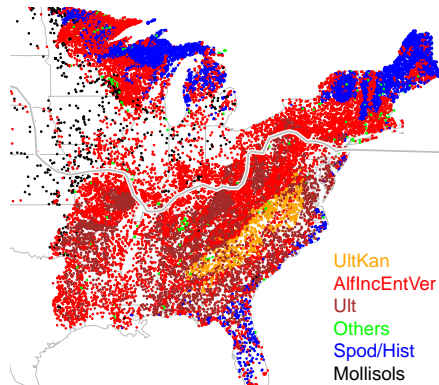


Figure S4: Soil types used as a multilevel factor.

**Surplus:** The hydrothermal surplus is the degree hours accumulated for months with a negative water balance (see Eq. S18)(Fig. S3a).

**Topography:** Slope and aspect are captured by three variables,

$$\mathbf{u}_i = \begin{cases} \sin(slope_i) \\ \sin(slope_i) & \sin(aspect_i) \\ \sin(slope_i) & \cos(aspect_i) \end{cases} \quad (S20)$$

There is a main effect of slope, but there can be no main effect of aspect independent of slope. If the first coefficient is not different from zero, then slope does not affect the response. If both of the last two coefficients are not different from zero then aspect does not affect the response (Clark 1990).

**Soils:** Soil types constitute a factor with 5 levels, Entisol-Vertisol, Mollisol, Spodosol-Histosol, Ultisol-Udults-Kanhapludults (Piedmont Plateau soils), and remaining Ultisols. Data were extracted from the State Soil Geographic (STATSGO) database (Fig. S4)). For each FIA plot, soil types are categorized based on the first to third soil taxonomy orders. To limit the number of levels in the model similar first order classes were merged into the categories listed above.

*Entisol* -poor profile development

*Vertisol* -plastic clays

*Mollisol* -former prairie rich soils

*Spodosol*, *Histosol* -moist, organic, nutrient-poor, low pH

*Ultisol*, *Udults-Kanhapludults* -clay rich

**Moisture index:** A FIA soil class variable related to moisture (Fig. S3d).

The predictors are not highly correlated with one another, and variance inflation factors are well below levels that would cause concern (i 3).

## 5.2 EMP

The EMP data used in this analysis are publicly available at <http://www.earthmicrobiome.org>.

## 6 Additional references

- 
- Clark, J.S., L. Iverson, C. W. Woodall, C. D. Allen, D. M. Bell, D. Bragg, A. D’Amato, F. W. Davis, M. Hersh, I. Ibanez, S. T. Jackson, S. Matthews, N. Pederson, M. Peters, M. W. Schwartz, K. Waring, and N. E. Zimmermann. 2015. The impacts of increasing drought on forest dynamics, structure, diversity, and management. In J. Vose, C. Luce, and J.S. Clark (eds). The National Assessment of Drought Impacts on Forests, in press.
- Sun, D. and J.O. Berger. 2006. Objective priors for the multivariate normal model. Proc. Valencia / ISBA 8th World Meeting on Bayesian Statistics.
- Yang, R. and Berger, J. O. 1994, Estimation of a covariance matrix using the reference prior. Annals of Statistics 22, 1195-1211.