The Annals of Applied Statistics
2019, Vol. 13, No. 4, 2363–2388
https://doi.org/10.1214/19-AOAS1281

PREDICTING PALEOCLIMATE FROM COMPOSITIONAL DATA USING MULTIVARIATE GAUSSIAN PROCESS INVERSE PREDICTION[1]

BY JOHN R. TIPTON*, MEVIN B. HOOTEN[†,‡], CONNOR NOLAN[§], ROBERT K. BOOTH[¶] AND JASON MCLACHLAN[‖]

University of Arkansas, Colorado State University[†], U. S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit[‡], University of Arizona[§], Lehigh University[¶] and University of Notre Dame[‖]*

Multivariate compositional count data arise in many applications including ecology, microbiology, genetics and paleoclimate. A frequent question in the analysis of multivariate compositional count data is what underlying values of a covariate(s) give rise to the observed composition. Learning the relationship between covariates and the compositional count allows for inverse prediction of unobserved covariates given compositional count observations. Gaussian processes provide a flexible framework for modeling functional responses with respect to a covariate without assuming a functional form. Many scientific disciplines use Gaussian process approximations to improve prediction and make inference on latent processes and parameters. When prediction is desired on unobserved covariates given realizations of the response variable, this is called inverse prediction. Because inverse prediction is often mathematically and computationally challenging, predicting unobserved covariates often requires fitting models that are different from the hypothesized generative model. We present a novel computational framework that allows for efficient inverse prediction using a Gaussian process approximation to generative models. Our framework enables scientific learning about how the latent processes co-vary with respect to covariates while simultaneously providing predictions of missing covariates. The proposed framework is capable of efficiently exploring the high dimensional, multi-modal latent spaces that arise in the inverse problem. To demonstrate flexibility, we apply our method in a generalized linear model framework to predict latent climate states given multivariate count data. Based on cross-validation, our model has predictive skill competitive with current methods while simultaneously providing formal, statistical inference on the underlying community dynamics of the biological system previously not available.

Received March 2019.

[1]This research is based upon work carried out by the PalEON Project (paleonproject.org) with support from the National Science Foundation MacroSystems Biology program under grant no. DEB-1241856. Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Key words and phrases. Bayesian hierarchical models, predictive validation, model comparison, ecological functional response model.

1. Introduction. A variety of data are used as proxies for climate, including tree rings, ice cores and pollen, with each data source requiring specialized statistical methods. Therefore, improving statistical techniques for reconstructing paleoclimate from proxy data are necessary to better understand past climate. Many climate proxies are nonnegative multivariate observations that occur on the unit simplex and arise when there is interest in explaining the proportion of a total. These types of data are called compositional count data and examples include fungal assays (Grantham et al. (2015), Saucedo-García et al. (2014)), molecular sequence data of soil microbes (Lauber et al. (2009)) and paleoecological data (Booth (2008), Brewer, Jackson and Williams (2012), Haslett et al. (2006), Paciorek and McLachlan (2009), Parnell et al. (2015), Salter-Townshend and Haslett (2012)). In general, compositional data are $N \times d$ counts $\mathbf{Y}$ where, for $i = 1, \ldots, N$, $\mathbf{y}_i$ is the d-dimensional composition for observation i (a given location or core depth) and $\sum_{j=1}^{d} y_{ij} = M_i$ is the total count of all species for observation i. When M_i is not informative of the total abundance of the composition, these data are called compositional count data and are one of the most common sources of paleoclimate proxy data. The observed composition at site i is assumed to depend on a set of covariates $\mathbf{x}_i$ that are only observed in the modern period and must be predicted for the past (see Nolan et al. (2019) for an example). In this manuscript, we develop novel statistical methods for generating predictions of unobserved climate $\mathbf{x}_i$ from observed compositional count data $\mathbf{y}_i$.

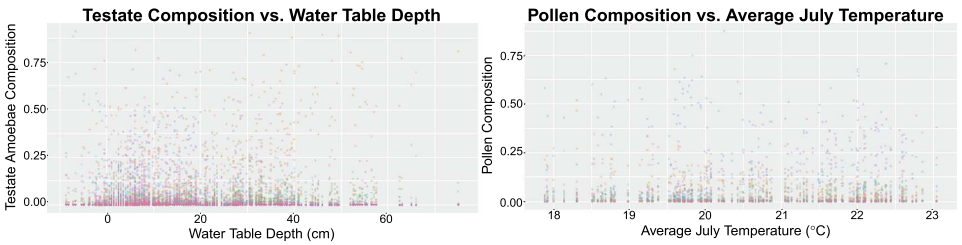
We present a new probabilistic model framework for paleoclimate reconstruction using compositional count data and apply our model alongside currently used Bayesian and deterministic transfer function methods to two compositional count datasets. Both of the application datasets are further subdivided into two subsets: (1) the modern calibration dataset where both the species compositions and climate variables are known and the (2) reconstruction data where only the species compositions are observed. The goal of paleoclimate reconstructions is to use the modern calibration dataset to learn model parameters and then use these parameters to predict the missing climate variables using the reconstruction data. We focus entirely on model performance for the modern calibration data because cross-validation is only possible during the modern calibration period. Therefore, we focus on the modern data only, with the assumption that models that perform well on the modern data will generalize to the reconstruction data. The two application datasets we consider are the compositional count of testate amoebae species in bogs relative to water table depth (a proxy measurement of hydroclimate) and the compositional count of pollen grains in lake sediments relative to average July temperature.

1.1. *Testate amoeba data.* Testate amoeba are a group of protists characterized by the presence of a test (i.e., shell) which live in a range of habitats including soils, wetlands and peat bogs. Because testate amoebae leave behind a decay-resistant test when they die and the tests can be used to identify different species of amoebae, sediment cores can be used as a record of the historical distribution of

testate amoeba species. If the distribution of testate amoebae species is sensitive to environmental conditions, examination of the distribution of species through time can be used to infer past climate.

Many studies have demonstrated a sensitivity of testate amoebae from ombrotrophic peatlands to moisture conditions (i.e., water table depth) (Amesbury, Barber and Hughes (2012), Booth, Lamentowicz and Charman (2010), Charman (2007)). To reconstruct water table depth (in cm), we used a modern-era data set with 356 replicates of testate amoeba assemblages including 24 species along a water depth gradient (Booth (2008)). We assumed that the species compositions of testate amoebae at a given water table depth are more likely to be correlated with compositions at other bog locations with similar water table depths than nearby locations in the same bog with different water table depths. We also made the assumption that the response of testate amoebae species to water depth is fixed through time. Under these assumptions, we used the data shown in Figure 1(a) to examine the relationship between the distribution of testate amoebae and water table depth. While many studies have used testate amoebae to reconstruct water depth, much remains to be learned about the ecology of the testate amoeba communities. Important ecological questions include how the testate amoebae species respond to the underlying environment (water table depth) and how they interact.

1.2. *Pollen data*. The second data source we consider is a set of pollen counts collected at 152 sites in the Upper Midwestern United States from sediments dating to the time of European settlement. The counts of pollen were identified to genus level and then grouped based on shared ecological characteristics into 16 categories. Pollen data have been studied extensively with many reconstructions of species compositions on the landscape (Dawson et al. (2016), Paciorek and McLachlan (2009)) and reconstructions of paleoclimate variables from pollen data (Wahl, Diaz and Ohlwein (2012), Williams and Shuman (2008), Haslett et al. (2006)). We are interested in predicting temperature using the Parameter elevation



(a) Testate amoeba data. (b) Pollen data.

FIG. 1. *Observed data with environmental covariate along the x-axis and species counts on the y-axis. Each color represents one of the d observed species. Figure* (*a*) *shows the testate amoeba data and Figure* (*b*) *shows the pollen data.*

Regression on Independent Slopes Model (PRISM) 30 year normal July temperature product in °C (PRISM Climate Group, Oregon State University) as a covariate. Figure 1(b) shows the relationship between species abundances and average July temperature. We made the assumption that there is no spatial correlation in pollen counts conditional on July temperature, although the model could be expanded in future work to incorporate spatially correlated random effects.

2. Model. We begin with a brief introduction to the currently used deterministic transfer function methods for paleoclimate reconstruction using compositional data. After introducing the most commonly used methods, we address the shortcomings of these methods to motivate the development of new probabilistic models. Next, we describe other Bayesian methods for paleoclimate reconstruction and describe the assumptions of these models that lead to the development of our model framework. We conclude this section by introducing the novel model framework and discuss how it solves many of the issues with previous methods.

2.1. *Transfer function methods*. The most widely used methods used in reconstructing climate from compositional data are generally known as "transfer functions." The three transfer function methods we consider are weighted averaging (WA), the modern analog technique (MAT) and maximum likelihood response curves (MLRC). WA, MAT and MLRC are methods that do not specify a joint likelihood for the data. WA generates reconstruction predictions by first learning parameters from the calibration data then applying those parameters to the reconstruction data. Learning model parameters from the calibration data occurs in three steps. First, WA estimates a nonparametric bootstrap distribution of "optimum" climate giving the prediction μ_j for species j as $\hat{\mu}_j = (\sum_{i=1}^n y_{ij}x_i)/(\sum_{i=1}^n y_{ij})$ using the modern calibration data, where y_{ij} is the observed proportion of species j in sample i and x_i is the climate variable for sample i. Second, out-of-bootstrap-sample estimates of climate $\hat{x}_i$ for sample i are generated by taking a weighted sum of the species "optima" $\hat{x}_i = \frac{\sum_{j=1}^d y_{ij}\hat{\mu}_j}{\sum_{i=1}^n y_{ij}}$ (Birks et al. (1990), ter Braak and van Dam (1989)). The WA method induces a strong shrinkage effect because the prediction $\hat{x}_i$ is a mean of means. The third step in generating the calibration model corrects for this shrinkage. The shrinkage to the mean is corrected by boosting the $\hat{x}_i$s with either a linear or spline regression between out-of-bootstrap-sample x_i and $\hat{x}_i$ resulting in a final model that is "de-shrunk" (Birks and Simpson (2013), Schapire (1990)). The most common implementations of WA in software then summarize the bootstrap sample with the bootstrap mean and variance to generate predictions under the assumption of a normal distribution (throwing away the information in the bootstrap sample). For samples $\tilde{i}$ in the reconstruction data, predictions of paleoclimate $\hat{x}_{\tilde{i}}$ are generated using $\hat{x}_{\tilde{i}} = \frac{\sum_{j=1}^d y_{\tilde{i}j}\hat{\mu}_j}{\sum_{\tilde{i}=1}^n y_{\tilde{i}j}}$ with uncertainties estimated from the normal approximation to the bootstrap distribution. The uncertainty for the paleoclimate prediction can be broken down into two sources: (1) the uncertainty

from the individual sample and (2) the average prediction error over the dataset. In practice, the average prediction error over the dataset dominates the uncertainty (often 95%+ of the total uncertainty in our experience). Therefore, the frequentist confidence interval widths are almost independent of the observed composition.

MAT is commonly known in the statistics and machine learning literature as k nearest neighbors. For each of the $\tilde{i}$ samples in the reconstruction data, a pseudo-distance is calculated to each of the $i = 1, \ldots, n$ observations in the calibration data. There are a variety of possible metrics, but the most commonly used in the literature is square chord distance (Overpeck, Webb and Prentice (1985)). The prediction $\hat{x}_{\tilde{i}}$ for samples in the reconstruction data is a (potentially weighted) average of the covariate of interest for the k nearest samples in the calibration data. Confidence intervals for the MAT prediction is generated by bootstrapping the calibration data with uncertainties estimated using out-of-bootstrap-sample uncertainties and assuming a normal error distribution.

MLRC fits a curve to each marginal species component of the composition then takes a weighted average of these curves to generate a prediction. The idea behind MLRC is that, at certain ranges of the climate space, different species would have higher/lower abundance. In implementation, a logistic nonlinear regression is fit to the presence/absence of each species marginally and the model predictions are made by finding the covariate value that gives rise to the highest probability of the observed composition. The simpler WA was developed to borrow the ideas of a functional response of the species to climate by reducing the computational and numerical challenges of MLRC (Birks et al. (1990), ter Braak and van Dam (1989)).

The transfer functions described above have been successful in reconstructing site-level point estimates of paleoclimate; however, transfer function methods have shortcomings. First, the transfer function methods estimate uncertainty using out-of-sample-bootstrap prediction errors that result in nearly uniform estimates of uncertainty for the reconstruction samples. This is not a desirable property because the the sample sizes used to generate the compositions change and the uncertainty estimates do not reflect the sampling intensity. In addition, there are compositions where the unobserved covariate can be predicted with higher precision than other samples based on domain knowledge, whereas current transfer functions methods are not capable of doing this. A third shortcoming is the sensitivity to compositions with zero counts; species with zero compositions are removed from the calibration dataset but sometimes the reconstruction datasets can have nonnegligible counts of these species. Removing these potentially meaningful species from the analysis could have an effect on the estimation process. Finally, both WA and MLRC model the response of each species marginally, ignoring the co-dependence and covariance among the composition.

In addition, the transfer function methods each have specific issues. In WA, species that are in high abundance dominate the prediction and the prediction is less sensitive to less abundant species. For MAT, there is no clear consensus on

what similarity metric is best and the final reconstruction might be sensitive to this choice. Under the "no-analog" setting (Overpeck, Webb and Webb III (1992) and Jackson and Williams (2004)) where the set of compositions are different between calibration and reconstruction datasets (i.e., there are no good "analogs" of the composition in the reconstruction data present in the calibration data because the distributions of species compositions are different), methods like WA and MAT show degraded predictive performance (Appendix S5, Tipton et al. (2019)). What makes the no-analog problem particularly nefarious is that one only uses the modern calibration data to evaluate model predictive performance leading to overly optimistic results when predictions are made using the reconstruction data. MLRC has been not widely used due to poor predictive performance and large confidence interval estimates—we include MLRC in our comparison as it is a likelihood-based method that serves as a methodological inspiration for our model.

Others have introduced Bayesian hierarchical models for reconstruction of paleoclimate using compositional data that overcome some of the issues with the transfer function methods. For instance, Vasko, Toivonen and Korhola (2000) introduced a Bayesian hierarchical model framework (BUMMER) to model compositional count data. BUMMER assumes that the marginal response of each species to climate follows a symmetric, unimodal Gaussian response curve. The BUMMER model uses a Dirichlet-multinomial likelihood (1) and assumes the latent random effect $\log(\alpha_{ij})$ for observation i of species j in (2) has the form of a Gaussian kernel

$$\log(\alpha_{ij}) = \exp\left(a_j - \frac{(b_j - x_i)^2}{2c_j^2}\right),$$

where a_j is an offset that models baseline abundance, b_j models the mode of the symmetric, unimodal response and c_j^2 is a parameter that models the spread of the functional response. In practice the data are often functional types or operational taxonomic units (OTUs) that combine many different species, each having one or more optimum responses that might respond asymmetrically to climate. Bhattacharya (2006) used a Dirichlet process mixture of unimodal Gaussian-shaped curves to model the response of each species to climate; however, the Dirichlet process approach is computationally demanding and is difficult to scale to large numbers of species. We will demonstrate that our model framework produces predictions that have similar skill to BUMMER when the symmetric, unimodal assumption is valid while greatly outperforming BUMMER predictive skill when the symmetric, unimodal assumption is violated.

We introduce a novel reconstruction framework that we call the multivariate Gaussian process model (MVGP) in the next section. In constructing the model, we show how the MVGP model addresses the criticisms of the models described above. Another strength of the MVGP model is the ability to estimate the marginal response curves for each species as well as the correlations among the functional

responses between species. Modeling inter-species correlations allows for further learning about the underlying ecology of the system as well as provides data-driven guidance on which taxa to combine together based on similar functional responses.

2.2. *Compositional data model.* Recall that compositional count data are $N \times d$ counts $\mathbf{Y}$ where, for $i = 1, \ldots, N$, $\mathbf{y}_i$ is the d-dimensional composition for observation i (a given location or core depth) and $\sum_{j=1}^{d} y_{ij} = M_i$ is the total count of all species for observation i. Because the data are counts, we specify the data model

$$\mathbf{y}_i \sim \text{Multinomial}(M_i, \mathbf{p}_i),$$

where the d-dimensional vector $\mathbf{p}_i \equiv (p_{i,1}, \ldots, p_{i,d})'$ represents the probabilities of sampling species j at location i under the constraint $\sum_{j=1}^{d} p_{ij} = 1$.

To ensure identifiability in multinomial models, one often assumes a reference category and fixes the value of the reference in the latent space. A consequence of including a reference category is that the random process $\mathbf{p}_i$ lives in a $(d-1)$-dimensional space. Because we are interested in allowing each species to have its own marginal response to the covariate while accounting for how these responses co-vary, we do not include a fixed reference category. In addition, visual inspection of the data and preliminary model fits suggest the compositional data are overdispersed. Therefore, we assign a mixing distribution for the probabilities $\mathbf{p}_i$ using the Dirichlet distribution, which allows modeling of each of the d functional responses to the covariates while adding a mechanism for overdispersion.

The Dirichlet distribution is commonly used to model distributions with sum-to-one constraints, including probabilities. We assume

$$\mathbf{p}_i \sim \text{Dirichlet}(\boldsymbol{\alpha}_i),$$

where the $\alpha_{ij} > 0$. Because the $\boldsymbol{\alpha}_i$ do not have a sum-to-one constraint, we model each of the d species directly using an appropriate link function. The choice of a conjugate prior model for $\mathbf{p}_i$ allows us to integrate out the latent variable $\mathbf{p}_i$, giving rise to the Dirichlet-multinomial distribution

$$\begin{aligned} \mathbf{y}_i &\sim \text{Dirichlet-Multinomial}(M_i, \boldsymbol{\alpha}_i) \\ &= \int \text{Multinomial}(\mathbf{y}_i | M_i, \mathbf{p}_i) \, \text{Dirichlet}(\mathbf{p}_i | \boldsymbol{\alpha}_i) \, d\mathbf{p}_i, \end{aligned} \tag{1}$$

where, for $i = 1, \ldots, N$ independent observations of the d-dimensional vector $\mathbf{y}_i$ of counts, the data model is

$$[\mathbf{y}_i | \boldsymbol{\alpha}_i] = \frac{\Gamma(\sum_{j=1}^{d} \alpha_{ij})}{\Gamma(M_i + \sum_{j=1}^{d} \alpha_{ij})} \prod_{j=1}^{d} \frac{\Gamma(y_{ij} + \alpha_{ij})}{\Gamma(\alpha_{ij})}.$$

To enforce the positive support of $\boldsymbol{\alpha}_i$, we use the log link function $\log(\boldsymbol{\alpha}_i) = \boldsymbol{\mu} + \boldsymbol{\zeta}_i + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{\zeta}_i$ is a vector of random effects that is conditional on the value of the climate covariate and $\boldsymbol{\varepsilon}_i$ is a vector of Gaussian error that accounts for overdispersion and is uncorrelated with $\boldsymbol{\zeta}_i$.

2.3. *Gaussian process model.* The latent functional responses for the correlated random effect $\boldsymbol{\zeta}$ are modeled using Gaussian processes. For $i = 1, \ldots, N$ replicates of the d-dimensional multivariate latent response $\log(\boldsymbol{\alpha}_i) \equiv (\log(\alpha_{i1}), \ldots, \log(\alpha_{id}))'$, we define the correlated multivariate Gaussian processes with nugget as

$$\log(\boldsymbol{\alpha}_i) = \boldsymbol{\mu} + \boldsymbol{\zeta}_i + \boldsymbol{\varepsilon}_i, \tag{2}$$

where $\boldsymbol{\mu}$ is a vector of means and/or fixed effects. For each observation $i = 1, \ldots, N$, the marginal distribution of the random effect $\boldsymbol{\zeta}_i \equiv (\zeta_{i1}, \ldots, \zeta_{id})'$ is mean zero Gaussian with covariance matrix $\boldsymbol{\Sigma}$ that accounts for the relationships among the observations (inter-species correlations). An important scientific question is how the d functional responses co-vary across covariate space. For example, when modeling distributions of species, similar functional responses to a covariate may indicate existence of a biologically relevant functional group. Thus, inference on $\boldsymbol{\Sigma}$ allows for formal learning about the underlying ecological relationships. We assume $\boldsymbol{\varepsilon}_i \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ is an independent and identically distributed Gaussian error vector that accounts for overdispersion in the data not explained by the Dirichlet mixture of multinomial distributions and can be added or removed from the model as necessary because it is not required for the generalized linear model.

For each dimension $j = 1, \ldots, d$, the marginal random effect $\boldsymbol{\zeta}_{1:N,j} \equiv (\zeta_{1,j}, \ldots, \zeta_{N,j})'$ is a Gaussian process with

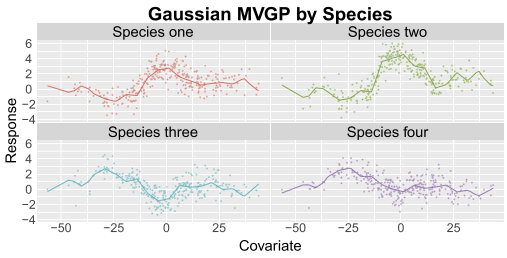
$$\boldsymbol{\zeta}_{1:N,j} \sim \mathrm{N}\big(\mathbf{0}, \mathbf{C}(\mathbf{X}, \boldsymbol{\theta})\big), \tag{3}$$

where $\mathbf{C}(\mathbf{X}, \boldsymbol{\theta})$ is a correlation matrix with i, i'th entry $\mathbf{C}_{i,i'} = c(\mathbf{x}_i, \mathbf{x}_{i'}, \boldsymbol{\theta})$ representing the latent correlation between observations at covariate values $\mathbf{x}_i$ and $\mathbf{x}_{i'}$, where $\mathbf{x}_i$ is the ith row of the $N \times q$ matrix of covariates $\mathbf{X}$. Hence, we assume the joint distribution of the random effects is $\boldsymbol{\zeta} \equiv (\boldsymbol{\zeta_1}', \ldots, \boldsymbol{\zeta}'_N)' \sim \mathrm{N}(\mathbf{0}, \mathbf{C}(\mathbf{X}, \boldsymbol{\theta}) \otimes \boldsymbol{\Sigma})$. Each of the d marginal Gaussian processes have unit variance for identifiability in the separable Kronecker product. The model can be generalized by allowing the covariance structure to be nonseparable, but such extensions are not explored in this manuscript.

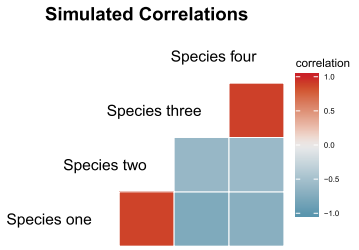
A common class of correlation functions is the Matérn class, of which the Gaussian and exponential correlation functions are examples (Stein (1999)). The Matérn correlation function is given by

$$c(\mathbf{x}_i, \mathbf{x}_{i'}, \boldsymbol{\theta}) = \frac{1}{\Gamma(\nu) s^{\nu-1}} \left(\frac{2\delta_{ii'}\sqrt{\nu}}{\rho} \right)^{\nu} \mathcal{K}_\nu \left(\frac{2\delta_{ii'}\sqrt{\nu}}{\rho} \right),$$

where $\delta_{ii'}$ is the Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_{i'}$, $\boldsymbol{\theta} = (\rho, \nu)'$, ρ is a spatial range parameter, and $\mathcal{K}_\nu(\cdot)$ is the modified Bessel function of the second kind with smoothness (differentiability) determined by the parameter $\nu > 0$. By letting $\nu \to \infty$, the Matérn correlation function converges to the Gaussian correlation



(a) Simulated Gaussian process response.



(b) Simulated correlations among species.

FIG. 2. *Simulation of correlated Gaussian processes for the log-scale random effect. Notice in the left figure that species one and two show a visually similar functional response to each other with respect to the covariate value, as do species three and four. The correlations are quantified by the correlation matrix on the right showing that species one and two have a positively correlated functional response, which is seen by the similarity in functional shape of the figure on the left. Likewise, species three and four have a similar functional shape and strong positive correlation. The other pairwise comparisons have different functional shapes and associated negative correlations.*

function, and setting $\nu = 0.5$ results in the exponential correlation function. For the examples in this manuscript, we use the exponential correlation function

$$c(\mathbf{x}_i, \mathbf{x}_{i'}, \theta) = \exp\left\{-\frac{\delta_{ii'}}{\rho}\right\},$$

where $\theta = \rho$, although the framework accommodates arbitrary correlation functions in general.

An example realization of a four-dimensional correlated Gaussian process on the log-scale of the latent random effect (Figure 2(a)) shows how the log-scale random effect varies in covariate space while allowing for the different dimensions (species) of the process to co-vary. The flexible nature of the multivariate Gaussian process allows for statistical learning about both the latent functional relationships for each species and how the functional responses interact among the species. For example, Figure 2(a) shows how species one and two co-vary in their respective functional responses due to a strong positive correlation (Figure 2(b)). We can see the influence of the correlation on the Gaussian process realizations where the first and second Gaussian processes (third and fourth) have very similar shapes while the other pairwise combinations have different shapes. The similar shapes between processes one and two (three and four) are because the processes are highly correlated with pairwise correlation 0.87 (0.88). The pairwise correlations of processes one and three, one and four, two and three, and two and four are -0.76, -0.6, -0.7 and -0.57, respectively. These negative correlations indicate that these species have a different functional response to the underlying covariate relative to each other. In the ecological setting, a positive correlation in functional response could represent a relevant functional group while a negative functional correlation could represent inhibition or competition (Morales-Castilla et al. (2015), Warton et al. (2015)).

We rewrite $\boldsymbol{\zeta}$ to improve computational efficiency in our MCMC algorithm. Using the upper triangular Cholesky decomposition $\boldsymbol{\Sigma} = \mathbf{R}'\mathbf{R}$, define $\boldsymbol{\zeta} \equiv (\boldsymbol{\eta}_1'\mathbf{R}, \ldots, \boldsymbol{\eta}_N'\mathbf{R})'$. For $j = 1, \ldots, d$, we define

$$\boldsymbol{\eta}_{1:N,j} | c(\cdot, \cdot | \boldsymbol{\theta}) \sim \mathrm{N}\big(\mathbf{0}, \mathbf{C}(\mathbf{X}, \boldsymbol{\theta})\big)$$

as independent Gaussian processes and rewrite the latent process (2) on the log-scale as

$$\log(\boldsymbol{\alpha}_i) = \boldsymbol{\mu} + \mathbf{R}'\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i, \tag{4}$$

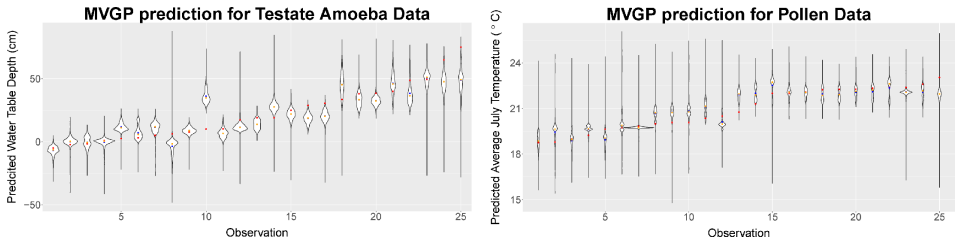
where the marginal distribution of $\boldsymbol{\eta}_i \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$ is multivariate standard Gaussian and the inter-species covariances are induced by the lower Cholesky matrix $\mathbf{R}$.

2.4. *Prior model.*

2.4.1. *Prior on the missing covariates.* To complete the model specification, we assign prior distributions to the remaining parameters. Because we are constructing a model for prediction of the missing, unobserved covariates given the observed covariates, we assign priors for the set of possible covariate values. For spatial and temporal problems, the support of $\mathbf{X}$ can be restricted to the domain of interest. For predictions in covariate space, the support of $\mathbf{X}$ should assign positive probability over the range of possible covariate values. We assign exchangeable priors on each of the $\tilde{i} = 1, \ldots, \tilde{N}$ unobserved covariate values

$$\mathbf{x}_{\tilde{i}} \sim \mathrm{N}(\boldsymbol{\mu}_X, 1.5\boldsymbol{\Sigma}_X), \tag{5}$$

with $\boldsymbol{\mu}_X$ and $\boldsymbol{\Sigma}_X$ chosen as the sample mean and covariance of the set of observed covariates. While this prior might not be fully Bayesian (it uses the observed covariates as a prior for the unobserved covariates), we argue that, a priori, it is reasonable to assume the past values of the covariate are similar to the current values with additional variation in the past relative to current values. Other prior specifications are possible as long as the prior is Gaussian (e.g., a correlated Gaussian spatio-temporal process) because the algorithm defined in Section 2.6 leverages a Gaussian prior distribution to efficiently sample the missing covariates. We assumed the Gaussian process correlation function is isotropic, hence $c(\mathbf{x}_{\tilde{i}}, \cdot, \boldsymbol{\theta})$ depends on $\mathbf{x}_{\tilde{i}}$ only through the set of pairwise distances to all of the other $\mathbf{X}$ (both observed and unobserved). Thus, given an estimate for the distance from a location, there are possibly many potential covariate values that can have posterior probability mass, giving rise to multiple modes in the posterior (i.e., the correlation function is a many-to-one function from which we aim to generate inverse predictions). The inclusion of multiple correlated Gaussian processes helps to constrain the multimodal estimation but does not fully resolve the multimodal problem when signal to noise ratio in the data model is low. The problem is also exacerbated by a non-Gaussian likelihood associated with the compositional data.



(a) Held-out testate amoeba data. (b) Held-out pollen data.

FIG. 3. *95% predictive distributions for* 25 *held-out observations from the testate amoeba data in Figure* (*a*) *and the pollen data in Figure* (*b*) *sorted to be increasing in the held-out covariate. The dots are at the held-out true values and the violin plots represent posterior distributions. Notice that the prediction distributions show evidence of the multimodality induced by the symmetry of the correlation function and the MCMC algorithm is capable of efficiently exploring the multiple modes. In addition, the MVGP model produces posterior distributions that have varying width for different samples.*

An example of the multimodal posterior distribution of the covariate is shown in Figures 3(a) and 3(b). The multimodality in the conditional posterior of $\mathbf{x}_{\tilde{i}}$ presents a challenge which requires an MCMC sampling algorithm that is capable of exploring multiple modes efficiently (hence the Gaussian prior (5); see Section 2.5 and Section 2.6 for details).

2.4.2. *Prior on the covariance* $\boldsymbol{\Sigma}$. The canonical inverse-Wishart prior for the covariance matrix $\boldsymbol{\Sigma}$ is often used for convenience because the inverse-Wishart is conjugate to a Gaussian likelihood, but the prior has some well-known shortcomings (Barnard, McCulloch and Meng (2000), O'Malley and Zaslavsky (2008)). In our case, the likelihood is non-Gaussian and we do not have the computational benefits of a conjugate prior. For these reasons, we explore an alternative prior on the covariance. We use a Cholesky decomposition $\boldsymbol{\Sigma} = \mathbf{R}'\mathbf{R}$ to induce the covariance that is vague and places marginal prior probability nearly uniformly over the space of possible correlations while assuming the marginal variances are independent of the correlations. We follow the separation strategy introduced in Barnard, McCulloch and Meng (2000) and model the marginal covariance $\boldsymbol{\Sigma} \equiv \operatorname{diag}(\boldsymbol{\tau})\boldsymbol{\Omega}\operatorname{diag}(\boldsymbol{\tau})$ using a diagonal matrix $\operatorname{diag}(\boldsymbol{\tau})$ with standard deviations $\boldsymbol{\tau} = (\tau_1, , \ldots, \tau_d)'$ on the diagonal, and $\boldsymbol{\Omega}$ is an arbitrary positive definite correlation matrix. We model the upper diagonal Cholesky decomposition $\mathbf{R}_\Omega$ of the correlation matrix $\boldsymbol{\Omega} = \mathbf{R}_\Omega'\mathbf{R}_\Omega$ directly, reducing computational cost by avoiding computation of the Cholesky decomposition (Pourahmadi (1999, 2000), Smith and Kohn (2002)). Thus, we place a prior directly on the upper diagonal Cholesky matrix $\mathbf{R} = \mathbf{R}_\Omega \operatorname{diag}(\boldsymbol{\tau})$ by modeling $\mathbf{R}_\Omega$ and $\boldsymbol{\tau}$.

Following Lewandowski, Kurowicka and Joe (2009) and the Stan Development Team (2016), we construct a prior for the Cholesky decomposition

of $\mathbf{R}_\Omega$ with upper diagonal elements constructed from $B = \binom{d}{2}$ random variables $\boldsymbol{\phi} \equiv (\phi_1, \ldots, \phi_B)'$ whose marginal support is the interval $(-1, 1)$. We model the ϕ_b as independent beta random variables where, for $b = 1, \ldots, B$, $\phi_b \sim 2\,\text{Beta}(\psi_b, \psi_b) - 1$ is a beta distribution scaled to the interval $(-1, 1)$. Lewandowski, Kurowicka and Joe (2009) call ϕ_b partial correlations in a regular vine and arrange $\boldsymbol{\phi}$ in an upper triangular matrix

$$\boldsymbol{\Phi} \equiv \begin{pmatrix} 0 & \phi_1 & \phi_2 & \cdots & \phi_{B-d+2} \\ 0 & 0 & \phi_3 & \cdots & \phi_{B-d+3} \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \phi_B \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

where the i, jth elements of $\boldsymbol{\Phi}$ (when $i < j$) are interpreted as the correlation between the projections of the ith and jth random variables on the plane orthogonal to all variables with row index i' less than i. From the partial correlations $\boldsymbol{\phi}$, we construct the Cholesky factor $\mathbf{R}_\Omega$ using the recursive relationship (Stan Development Team (2016))

$$R_{\Omega i,j} = \begin{cases} 0 & \text{if } i > j, \\ 1 & \text{if } i = j = 1, \\ \displaystyle\prod_{i'=1}^{i-1} \sqrt{1 - \Phi_{i',j}^2} & \text{if } 1 < i = j, \\ \Phi_{i,j} & \text{if } 1 = i < j, \quad \text{and} \\ \Phi_{i,j} \displaystyle\prod_{i'=1}^{i-1} \sqrt{1 - \Phi_{i',j}^2} & \text{if } 1 < i < j, \end{cases}$$

where $R_{\Omega i,j}$ and $\Phi_{i,j}$ are the i, jth elements of $\mathbf{R}_\Omega$ and $\boldsymbol{\Phi}$, respectively.

To complete the prior on the separation model for $\boldsymbol{\Sigma}$, we follow the recommendations in Gelman (2006) and assign independent half-Cauchy priors on the standard deviations $\boldsymbol{\tau}$, where for $j = 1, \ldots, d$, $\tau_j \sim \text{half-Cauchy}(0, s_j)$. To sample from the half-Cauchy priors efficiently, we use the result from Armagan, Dunson and Clyde (2011) where, for $j = 1, \ldots, d$, $\tau_j^2 \sim \text{Gamma}(\frac{1}{2}, \lambda_j)$ with mixing parameter $\lambda_j \sim \text{Gamma}(\frac{1}{2}, s_j^2)$ inducing a half-Cauchy $(0, s_j)$ distribution on the standard deviation τ_j. The prior on the uncorrelated residual error covariance $\boldsymbol{\Sigma}_\varepsilon$ is constructed using the same separation strategy as for $\boldsymbol{\Sigma}$. Depending on model performance in cross-validation in the generalized linear mixed model framework, the uncorrelated random effect $\boldsymbol{\varepsilon}_i$ can be included or dropped from the model.

2.5. *Computational considerations*. When the covariates $\mathbf{X}$ are fixed and known, parameter estimation for model (2) requires on the order of $O(dN_{\text{total}}^3)$[2] operations, with the computationally costly inversion and determinant of the $N_{\text{total}} \times N_{\text{total}}$ dimensional matrix $\mathbf{C}$ dominating the computation of the likelihood, where $N_{\text{total}} = N + \tilde{N}$ is the total number of locations. If N_{total} is large, inversion of this matrix alone is computationally challenging. For the inverse problem, where the goal is estimation of a set of $\tilde{N}$ unknown covariates $\mathbf{x}_{\tilde{i}}$, the computational cost is prohibitive for even small N_{total}, because each MCMC iteration (or optimization step) over the $\tilde{N}$ missing covariates is $O(dN_{\text{total}}^3)$ and needs to be repeated for each of the $\tilde{N}$ unobserved covariates resulting in the prohibitive computational cost of $O(\tilde{N}\, dN_{\text{total}}^3)$. Thus, there is need for a computationally tractable approach for estimation under a multivariate Gaussian process model. One approach is to use a rank one Cholesky update on the full rank correlation matrix, which updates the Cholesky of the correlation matrix $\mathbf{C}$ given a single row change in $\mathbf{X}$. The rank one Cholesky update has the much lower computational cost of $O(dN_{\text{total}}^2)$ to update the unobserved covariates because this operation uses sums of outer products of vectors, but these operations must be repeated $\tilde{N}$ times resulting in total computational complexity of $O(\tilde{N}\, dN_{\text{total}}^2) + O(dN_{\text{total}}^3)$ and we found the algorithm numerically unstable and too computationally expensive (Seeger (2004)).

A number of approaches to dimension reduction for Gaussian process models have been proposed in the literature when the goal is traditional prediction (called Kriging in the geostatistical literature), with respective advantages and disadvantages. Lindgren, Rue and Lindström (2011) proposed approximating the Gaussian process with a Markov random field by transforming the data to a regular lattice. The Gaussian Markov Random Field approach requires the space over which the Gaussian process occurs to be known, which is not appropriate for our problem. It may be possible to allow for changing locations on the lattice, but the computational difficulty in implementing this method is substantial. Another approach for reducing computational cost in Gaussian process models is to approximate the likelihood in the spectral domain (Fuentes (2007), Paciorek (2007), Stein (1999)). Spectral methods are difficult to implement for the traditional prediction problem, let alone for inverse prediction, because they require expert tuning and require the space over which the random process occurs to be known. Another class of methods can be described as local approximations and express the joint likelihood as a product of conditional distributions. Examples of local likelihood methods include the nearest neighbor Gaussian process (Datta et al. (2016)) and the compositional likelihood introduced by Vecchia (1988). These methods suffer from issues in the

[2]Although in practical implementations, the limiting computational bottleneck is the calculation of the Cholesky decomposition of $\mathbf{C}$, which is of order $O(N_{\text{total}}^3/3)$ and must be repeated for each of the d Gaussian processes.

inverse prediction problem because the meaning of local is ill-defined when the values over which the Gaussian process occurs are unknown.

A final class of methods for approximating Gaussian processes are linear combinations of stochastic processes, including kernel convolutions, wavelet regression and other methods that can broadly be described as basis function expansions (Banerjee et al. (2008), Cressie and Johannesson (2008), Hefley et al. (2017), Higdon (2002), Kammann and Wand (2003), Nychka et al. (2015)). Basis function methods allow the random effect to be separate from the space over which the Gaussian process occurs by interpolating from a set of fixed locations, commonly referred to as knots. Instead of the location information residing in the correlation function, the information about location occurs in the basis expansion. Thus, any computation that depends on the unknown covariates requires matrix multiplication instead of matrix inversion. Although there are many different basis function models, we choose the predictive process (Banerjee et al. (2008)). The predictive process approximation to the Gaussian process has the properties that the parameters of the low-rank process have the same interpretation as the parameters of the full-rank process, and the predictive process is the best low-rank approximation of order ℓ in terms of minimizing Kullback–Leibler divergence (Csató (2002)). Another advantage of the predictive process is that one can implement our proposed model for any arbitrary positive definite correlation function, allowing for a broad, flexible class of Gaussian process models where inverse prediction is possible.

By assuming the Gaussian process can be represented by a low-rank predictive process approximation of order ℓ, we approximate the marginal model for the random effect (3) as

$$\begin{aligned}\boldsymbol{\eta}_{1:N,j} &\approx \mathbf{c}^{\star}(\mathbf{X}, \mathbf{X}^{\star}, \boldsymbol{\theta})\mathbf{C}^{\star}(\mathbf{X}^{\star}, \boldsymbol{\theta})^{-1}\boldsymbol{\eta}_j^{\star} \\ &= \mathbf{Z}(\mathbf{X}, \mathbf{X}^{\star}, \boldsymbol{\theta})\boldsymbol{\eta}_j^{\star},\end{aligned}$$

where $\mathbf{C}^{\star}(\mathbf{X}^{\star}, \boldsymbol{\theta})$ is the correlation matrix created by evaluating the correlation function $c(\mathbf{X}^{\star}, \mathbf{X}^{\star}, \boldsymbol{\theta})$ at the ℓ fixed knots $\mathbf{X}^{\star}$, $\mathbf{c}^{\star}(\mathbf{X}, \mathbf{X}^{\star}, \boldsymbol{\theta}) = c(\mathbf{X}, \mathbf{X}^{\star}, \boldsymbol{\theta})$ is the cross-correlation of the process at the covariate locations $\mathbf{X}$ and the knots $\mathbf{X}^{\star}$, $\mathbf{Z}(\mathbf{X}, \mathbf{X}^{\star}, \boldsymbol{\theta}) = \mathbf{c}^{\star}(\mathbf{X}, \mathbf{X}^{\star}, \boldsymbol{\theta})\mathbf{C}^{\star}(\mathbf{X}^{\star}, \mathbf{X}^{\star}, \boldsymbol{\theta})^{-1}$ is the basis expansion matrix that contains known and unknown covariate values, and the low-rank random process defined at the knots is the Gaussian predictive process

$$\boldsymbol{\eta}_j^{\star} \sim \mathrm{N}(\mathbf{0}, \mathbf{C}^{\star}(\mathbf{X}^{\star}, \mathbf{X}^{\star}, \boldsymbol{\theta})).$$

Most importantly for the inverse problem, the location information (known and unknown) is in the matrix $\mathbf{Z}(\mathbf{X}, \mathbf{X}^{\star}, \boldsymbol{\theta})$ rather than the correlation function. Thus, estimating the unknown Gaussian process hyperparameters requires the $O(d\ell^3) \ll O(dN_{\text{total}}^3)$ inversion of the reduced-rank matrix $\mathbf{C}^{\star}(\mathbf{X}^{\star}, \mathbf{X}^{\star}, \boldsymbol{\theta})$ once per MCMC iteration (this can be done using a cached Cholesky decomposition that is reused for each of the $\tilde{N}$ missing covariates) and evaluation of the likelihood for each

observation with an unknown covariate is now reduced from the $O(dN_{\text{total}}^3)$ matrix inversion for the full model to the $O(d\ell^2)$ vector-matrix multiplication $\mathbf{z}_i\boldsymbol{\eta}^\star$, where $\mathbf{z}_i$ is the ith row of $\mathbf{Z}(\mathbf{X}, \mathbf{X}^\star, \boldsymbol{\theta})$. Hence, we have reduced the computational cost from $O(\tilde{N}\, dN_{\text{total}}^3)$ to $O(d\,\ell^3) + O(\tilde{N}\, d\ell^2)$. The reduced computational cost for estimating unknown covariates means the predictive process approach is well suited for our model framework, despite the predictive processes' well-known shortcomings of producing estimates that are overly smooth at fine-scales (Finley et al. (2009), Stein (2014)).

Using the predictive process, we write our approximate Gaussian process random effect as

$$\boldsymbol{\zeta}_i^\star = \mathbf{R}'\big(\mathbf{z}_i\boldsymbol{\eta}^\star\big)' \tag{6}$$

and rewrite the latent random effect on the log-scale (4) as

$$\log(\boldsymbol{\alpha}_i) = \boldsymbol{\mu} + \boldsymbol{\zeta}_i^\star + \boldsymbol{\varepsilon}_i. \tag{7}$$

2.6. *Implementation.* The posterior we sample from using MCMC is

$$\begin{aligned}
&\big[\boldsymbol{\eta}^\star, \sigma^2, \lambda_\sigma, \tau^2, \lambda_\tau, \rho, \boldsymbol{\phi}, \{\mathbf{x}'\}_{\tilde{i}=1}^{\tilde{N}} | \{\mathbf{y}_i, i = 1, \ldots, N_{\text{total}}\}\big] \\
&\quad \propto \prod_{i=1}^{N_{\text{total}}} \big[\mathbf{y}_i|\boldsymbol{\eta}^\star, \sigma^2, \tau^2, \rho\big]\big[\boldsymbol{\eta}^\star|\tau^2, \rho\big]\big[\sigma^2|\lambda_{\sigma^2}\big]\big[\lambda_{\sigma^2}\big] \\
&\qquad \times \big[\tau^2|\lambda_{\tau^2}\big]\big[\lambda_{\tau^2}\big]\big[\rho\big]\big[\boldsymbol{\phi}\big] \prod_{\tilde{i}=1}^{\tilde{N}} \big[\mathbf{x}_{\tilde{i}}\big],
\end{aligned}$$

which is high dimensional and multimodal. Thus, we develop a MCMC algorithm that is fast, efficient and can explore a multimodal distribution effectively. The elliptical slice sampler is a highly efficient method for sampling from parameters whose prior distribution is multivariate Gaussian because the elliptical slice sampler has no tuning parameters, jointly updates parameters of interest, reduces random walk behavior and allows for exploration of multimodal full-conditionals (Murray, Adams and MacKay (2010)). We found the elliptical slice sampler to be highly effective in sampling the high-dimensional predictive process random effect $\boldsymbol{\eta}^\star$ and the multimodal unobserved missing covariates $\mathbf{x}_{\tilde{i}}$; elliptical slice sampling vastly outperformed importance sampling within Gibbs and Metropolis within Gibbs algorithms (Gelfand and Smith (1990)). For other parameters, we use an adaptive random walk Metropolis within Gibbs algorithm, with multivariate proposals on the log- or logit-scale if appropriate (Roberts and Rosenthal (2009)).

All the probabilistic models were fit using 200,000 MCMC iterations, discarding the first 50,000 iterations as burn-in and fixing the adaptive proposal distributions for the remaining 150,000 iterations. Each algorithm was run for four parallel chains, thinning every 150 iterations to reduce output file size resulting in 4000

posterior samples. Convergence was assessed using the Gelman–Rubin $\hat{R}$ statistic (Gelman, Rubin et al. (1992)). For the predictive process, we assign 30 knots evenly spaced over a range extending 1.5 standard deviations of the observed climate state beyond the minimum and maximum observed climate state. Based on repeated model fits, we found the inference obtained from the MVGP model to be insensitive to the number and location of knots.

To increase computational speed, we coded the algorithm in C++ using the `RcppArmadillo` package (Eddelbuettel and Sanderson (2014)) within the `R` computing environment (R Core Team (2016)). We fit competing models in `R` using the `rioja` package (Juggins (2015)).

3. Empirical model evaluation. We evaluate the performance of the model framework using a simulation study as well as a cross-validation experiment using the two representative compositional count datasets. For paleoclimate proxy databases, there are often one or two environmental covariates that are measured in common across different studies. Therefore, we focus on the predictive performance of the model framework to a single covariate variable in what follows. To explore the impact of model assumptions on predictive performance we consider two simulation experiments. For the first experiment, we simulated data from the BUMMER model. The second experiment simulated data under the MVGP model. For each of these datasets, we can compare the estimates of the parameters from the equivalent simulated model to demonstrate that each model (BUMMER and MVGP) is capable of recovering the underlying simulated parameters. In addition, we can compare the influence of the symmetric, unimodal assumption of BUMMER relative to MVGP in simulated data. The simulation-based experiments evaluate predictive performance using out-of-sample test data simulated from the model under consideration. We also estimate the missing covariates using the transfer-function methods and compare with predictions generated using the proposed model. For comparisons, we also fit a simplified generalized additive model (GAM) version of the MVGP model using a B-spline approximation to the latent functional response with a Dirichlet-multinomial likelihood.

For the cross-validation experiments, we evaluate model performance using the calibration testate amoeba and pollen data. Model performance is assessed by comparing predictive skill using k-fold cross validation for each of the candidate models. To balance computation time and to avoid model stability issues by holding-out too much data, we cross-validated over 12-fold hold-out data sets chosen at random. All of the empirical model experiments do not include the additional overdispersion term $\boldsymbol{\varepsilon}_i$ as inclusion of this term did not improve predictive skill in cross-validation.

3.1. *Model evaluation.* We evaluated the reconstructions using mean square prediction error (MSPE), mean absolute error (MAE), empirical 95% coverage for

either central Bayesian 95% credible intervals or 95% frequentist confidence intervals, and the continuous ranked probability score (CRPS) (Gneiting (2011)). One desirable property of a scoring rule is propriety. A proper scoring rule is one which, under expectation, selects the optimal predictive model. A strictly proper scoring rule is one which, under expectation, selects the optimal predictive model and no others. MSPE (MAE) has the advantage of being widely used, easy to understand and easy to implement, but is not a strictly proper scoring rule for arbitrary likelihoods. To understand why MSPE (MAE) is not strictly proper, one can envision two models that yield the same predictive mean but different predictive variances. The best model is the one with empirical coverage closest to the nominal rate, but MSPE (MAE) is unable to distinguish between the two predictions and is thus not strictly proper. However, MSPE (MAE) is proper because the best model has the lowest score, on average. To ensure that our scores are proper, we define the point forecast as the mean (median) of the predictive distribution when using MSPE (MAE). Instead of looking at Bayesian credible interval or frequentist confidence interval coverage and MSPE or MAE jointly, one can use a strictly proper scoring rule that integrates this idea formally. CRPS is a scoring rule that rewards predictions that are close to the true value while also rewarding proper estimation of uncertainty (Gneiting, Balabdaoui and Raftery (2007)). The CRPS is defined for the cumulative predictive distribution F and out-of-sample realization y_{oos} as

$$\text{CRPS}(F, y_{\text{oos}}) = \int_{-\infty}^{\infty} \big(F(y) - \mathbf{I}(y \geq y_{\text{oos}})\big)^2 \, dy.$$

When fitting a Bayesian model using sampling methods, including MCMC, one can approximate the CRPS using

$$\text{CRPS} = \frac{1}{2K^2} \sum_{k=1}^{K} \sum_{\kappa=1}^{K} |y^{(k)} - y^{(\kappa)}| - \frac{1}{K} \sum_{k=1}^{K} |y^{(k)} - y_{\text{oos}}|,$$

where $y^{(k)}$ and $y^{(\kappa)}$ are the kth and κth samples from the posterior predictive distribution. For the nonprobabilistic WA, MAT and MLRC methods that produce point forecasts, the CRPS score is equivalent to MAE.

3.2. *Results*. For the first simulated data experiment, the BUMMER model that assumes a symmetric, unimodal functional response of each species to climate was used to generate the data. Details about the simulation parameters can be found in Appendix S1 (Tipton et al. (2019)). Predictive scores in Table 1(a) show that, for data simulated under the BUMMER model, the BUMMER model slightly outperforms MVGP and the simpler GAM across all metrics and all the probabilistic methods outperform the transfer function methods. This is unsurprising as the BUMMER model was used to simulate the data and is the correct model; however, the GAM and MVGP models produce predictions with essentially equal skill. The transfer function methods fail to perform as well as the probabilistic models under the simulated BUMMER data.

TABLE 1

Results for predicting unobserved covariate values using simulated data. Smaller MSPE, MAE and CRPS values indicate better model performance. Central Bayesian credible interval and frequentist confidence interval coverage values closer to the nominal 95% credible interval indicate better model performance

(a) Simulated BUMMER data

	CRPS	MSPE	MAE	95% CI coverage
MVGP	0.6502	1.4008	0.9105	95.0000
GAM	0.6456	1.3683	0.9121	96.0000
BUMMER	0.6397	1.3480	0.9154	93.5000
WA	1.1143	1.8814	1.1143	98.0000
MAT	1.1057	1.9993	1.1057	98.5000
MLRC	1.6331	4.3291	1.6331	100.0000

(b) Simulated MVGP data

	CRPS	MSPE	MAE	95% CI coverage
MVGP	0.5162	1.2637	0.6793	96.0000
GAM	0.5394	1.3372	0.7113	96.5000
BUMMER	0.7198	1.7937	1.0057	96.0000
WA	1.0962	1.9467	1.0962	96.0000
MAT	0.9413	1.6814	0.9413	95.0000
MLRC	1.3245	3.3055	1.3245	95.0000

For the second simulation experiment, we simulated compositional count data from the MVGP model and generated predictions using the candidate models (Appendix S2, Tipton et al. (2019)). The results in Table 1(b) show the MVGP model performs best across all metrics with the simpler GAM model performing nearly as well. The BUMMER model shows decreased predictive skill because the symmetric, unimodal functional response assumption has been violated. The transfer function methods fail to perform as well as MVGP and GAM, with MAT outperforming WA and MLRC. The MVGP model estimated the parameters simulated from the MVGP model accurately, demonstrating that MVGP is useful for prediction and inference (Appendix S2, Tipton et al. (2019)). By providing good predictions and accurately estimating the simulated parameters in data similar to the observed data of interest, the MVGP framework is shown to be useful for both prediction and inference using the simulated data. From this experiment, we have shown that the BUMMER model shows degraded predictive performance when the assumption of a symmetric, unimodal functional response is violated.

We used 12-fold cross-validation to test model performance on the testate amoeba and pollen data to better explore the predictive performance of the candidate models. Details for the experiments can be found in Appendices S3 and

TABLE 2

Cross-validation scores for experiment three using the application data. Smaller MSPE, MAE and CRPS values indicate better model performance. Central Bayesian credible interval and frequentist confidence interval coverage values closer to the nominal 95% *credible interval indicate better model performance*

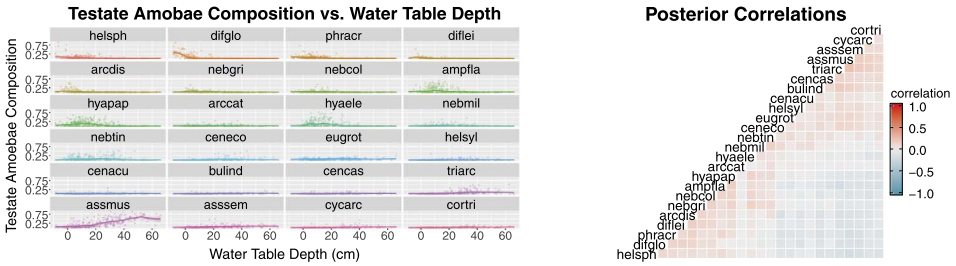
(a) Testate amoeba data

	CRPS	MSPE	MAE	95% CI coverage
MVGP	0.2959	0.2793	0.3897	78.3708
GAM	0.2966	0.2933	0.3943	79.7753
BUMMER	0.3073	0.3020	0.4077	75.8427
WA	0.4014	0.2924	0.4014	94.3820
MAT	0.3867	0.2340	0.3867	98.8764
MLRC	0.4088	0.4310	0.4088	92.6966

(b) Pollen data

	CRPS	MSPE	MAE	95% CI coverage
MVGP	0.2746	0.2178	0.3864	87.5000
GAM	0.2828	0.2119	0.3912	78.9474
BUMMER	0.2895	0.2485	0.4001	79.6053
WA	0.3711	0.2287	0.3711	94.0789
MAT	0.3486	0.1821	0.3486	100.0000
MLRC	0.3880	0.2722	0.3880	96.7105

S4 (Tipton et al. (2019))for the testate amoeba and pollen data, respectively. The cross-validation experiment results in Table 2(a) (testate amoeba data) and Table 2(a) (pollen data) demonstrate that the MVGP and GAM models generate the best probabilistic predictions, whereas MAT produced the best point predictions with MVGP, GAM and WA performing similarly on these metrics. For the CRPS metric MVGP, GAM and BUMMER are the best performing models in that order. The MAT produced confidence intervals with slightly higher than expected coverage; MVGP and GAM produced central 95% Bayesian credible intervals with lower than expected empirical coverage. In general, the cross-validated central Bayesian credible interval coverage for MVGP, GAM and BUMMER is low for the real data sets, perhaps due to additional overdispersion that is not accounted for in the data model. The MLRC method produced the worst predictions across both studies, but is included because MLRC is a functional response method like MVGP and GAM and is robust to the no-analog problem (Appendix S5, Tipton et al. (2019)). By being competitive with currently used predictive methods, the cross-validation experiment demonstrates the potential of the MVGP inverse prediction framework for paleoclimate reconstruction. The predictions shown in Figure 3(a) and 3(b) demonstrate that the model generates reasonable predictive distributions,

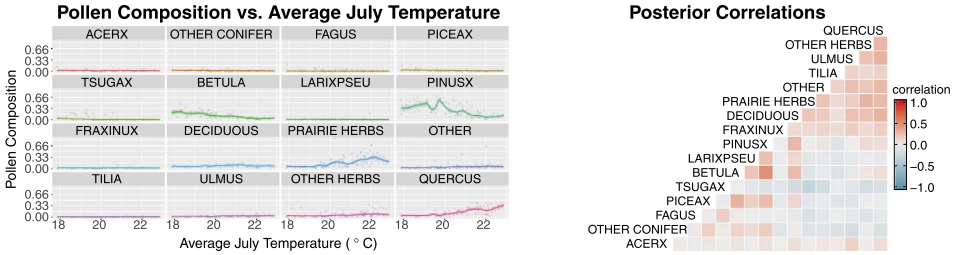


(a) Posterior mean response. (b) Posterior correlations.

FIG. 4. *Posterior model fit for the testate amoeba data. Figure* (*a*) *shows the fitted latent response of each species' contribution to the composition with respect to water table depth, sorted by increasing maximal species response to water table depth from top left to bottom right. Figure* (*b*) *shows the posterior mean pairwise correlations in response to water table depth among testate amoeba species, sorted by increasing maximal species response to water table depth from bottom left to top right. Red colors show positive correlations and blue colors show negative correlation.*

but the multimodality suggests that the posterior mean (or median) implied by use of MSPE (MAE) might not always a good description of the predictive distribution for MVGP and GAM—hence the decreased performance on the point prediction metrics of MSPE and MAE.

The MVGP, BUMMER and GAM models also allow for inference that is not available with the transfer function methods (WA, MAT and MLRC). The ability to make inference on the latent functional relationship between composition data and unobserved covariates is valuable. Figures 4(a) and 5(a) show the MVGP posterior mean response of each testate and pollen species to water table depth and average July temperature, respectively. Using these figures, researchers can make meaningful inference about the ecological niche that different species are exploiting.



(a) Posterior mean response. (b) Posterior correlations.

FIG. 5. *Posterior model fit for the pollen data. Figure* (*a*) *shows the fitted latent response of each species' contribution to the composition with respect to average July temperature, sorted by increasing maximal species response to temperature from top left to bottom right. Figure* (*b*) *shows the posterior mean pairwise correlations in response to average July temperature among species, sorted by increasing maximal species response to average July temperature from bottom left to top right. Red colors show positive correlations and blue colors show negative correlation.*

For example, in Figure 4(a) the species *Assulina muscorum* (assmus) is dominant in environments with water table depths deep below the peatland surface while the species *Hyalosphenia elegans* (hyaele) is more prevalent when water tables are near the surface. In Figure 5(a), there is a pronounced multimodal response of *Pinus* species to average July temperature. The *Pinus* taxa is a combination of different pine species that each have a different ecological niche and would be expected to have a multimodal functional response that would not be detected under the BUMMER model.

The posterior mean estimates of the correlations among functional responses shown in Figure 4(b) and Figure 5(b) provide insight into potential interactions among species. For example, some members of the same genus, such as *Nebula* (neb) and *Assulina* (ass), show positive correlations in their functional response to water depth. Not surprisingly though, there are a number of unrelated species that show high correlations, such as *Trigonopyxis arcula* (triarc) and *Assulina spp* (assmus, asssem), likely because they occupy similar niches with respect to surface moisture and other environmental conditions. This is supported in Figure 4(b) by the presence of positive correlations near the diagonal of the correlation matrix. The positive correlations near the diagonal in Figure 4(b) suggest species that have maximal responses that are near each other display a correlation in the functional response to water table depth. Importantly, species with similar energetic strategies, in particular mixotrophic species like *Hyalosphenia papilio* (hyapap) and *Heleopera sphagni* (helsph) that contain endosymbiotic zoochorellae also have high correlations, likely due to similar water-table depth tolerances as well as light requirements. Likewise Figure 5(b) shows the positive correlations of plant species with maximal responses at either the cool or hot end of the average July temperature gradient. For example, *Pinus* and *Betula* are highly correlated in their functional response, which is not surprising as these taxa frequently occur together. Species with high correlations could be clustered together in future statistical analysis to reduce the number of parameters that need to be estimated and might also save analysts time by reducing the need for taxonomic precision in some cases. These inferential questions cannot be answered by the transfer function techniques commonly used to model compositional data, demonstrating the value of the MVGP modeling approach.

4. Discussion. This work developed a flexible, novel model framework for prediction of unobserved climate from compositional count data. We have shown the MVGP model is capable of providing predictions that are of equal or superior skill to current methods depending on the properties of the underlying data. Although MVGP might not always be the most skilled model in cross-validation using the observed data, the MVGP model is robust to the issues commonly seen in using compositional count data to reconstruct climate including the "no-analog" setting. In "no-analog" data, the cross-validation scores for MVGP dominate the other models (Appendix S5, Tipton et al. (2019)). Thus, although MVGP does not

uniformly outperform WA and MAT in cross-validation, the empirical robustness of MVGP to the "no-analog" problem (Appendix S5, Tipton et al. (2019)) and performance on data simulated from the BUMMER model suggests that our model framework is more robust to data seen in practice.

We have shown that the MVGP model can be a useful inferential tool by fitting the model to simulated data and showing that the inference accurately reproduces simulated parameters. This is important because posterior inference that we obtain from the MVGP provides learning about the underlying processes that is not available in the transfer function methods WA and MAT while making fewer assumptions than the current Bayesian methodology of BUMMER. The learning about the correlations in functional responses can be used to validate the groupings of taxa into similar functional types and provide the basis for a data-driven method of grouping taxa based on similar functional responses. Using the learning of the underlying processes giving rise to the data, we can guide future model and dataset development to iteratively improve paleoclimate reconstructions.

The statistical and computational framework developed in this manuscript allowed us to fit a complex model to compositional count data in a reasonable amount of time. We presented an algorithm for predicting unobserved inputs into a Gaussian process model by resolving a computational bottleneck that allowed for use of the highly flexible Gaussian process functional form to be used for inverse prediction. We also implemented an MCMC sampling technique that can efficiently explore the multimodal posterior generated by the inverse prediction framework.

Finally, MVGP and the other probabilistic methods for climate reconstruction have greater flexibility in future development. It is natural to extend the probabilistic models to include temporal or spatial autocorrelation by assigning the unobserved covariates $\mathbf{X}$ a correlated prior. Possible autocorrelation structures can account for either continuous or discrete observations in space and time. In addition, because the likelihood is probabilistic, one can account for radio-dating uncertainty through weighting the likelihoods using an estimated age-depth model to more fully account for uncertainty. The impact of these changes is difficult to include in a cross-validation experiment but better incorporates domain knowledge of how the climate processes evolve in time and space. Given that the MVGP model is competitive in predictive skill with current modeling efforts, the ability to extend MVGP to correlated spatio-temporal reconstructions is a great benefit. Additional improvements in reconstruction skill could also be seen by combining MVGP with other proxy data sources to provide more precise inference of paleoclimate.

SUPPLEMENTARY MATERIAL

Code for replication of results and appendices (DOI: 10.1214/19-AOAS1281SUPP; .zip). The zip file includes detailed analyses in the appendices as well as

code and data for replication of the analyses presented in the manuscript and appendices.

REFERENCES

AMESBURY, M., BARBER, K. and HUGHES, P. (2012). The relationship of fine-resolution, multi-proxy palaeoclimate records to meteorological data at Fågelmossen, Värmland, Sweden and the implications for the debate on climate drivers of the peat-based record. *Quat. Int.* **268** 77–86.

ARMAGAN, A., DUNSON, D. B. and CLYDE, M. (2011). Generalized Beta mixtures of Gaussians. *Adv. Neural Inf. Process. Syst.* **24** 523–531.

BANERJEE, S., GELFAND, A. E., FINLEY, A. O. and SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 825–848. MR2523906

BARNARD, J., MCCULLOCH, R. and MENG, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statist. Sinica* **10** 1281–1311. MR1804544

BHATTACHARYA, S. (2006). A Bayesian semiparametric model for organism based environmental reconstruction. *Environmetrics* **17** 763–776. MR2252330

BIRKS, H. J. B. and SIMPSON, G. L. (2013). 'Diatoms and pH reconstruction' (1990) revisited. *J. Paleolimnol.* **49** 363–371.

BIRKS, H. J. B., LINE, J. M., JUGGINS, S., STEVENSON, A. C. and TER BRAAK, C. J. F. (1990). Diatoms and pH reconstruction. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **327** 263–278.

BOOTH, R. K. (2008). Testate amoebae as proxies for mean annual water-table depth in Sphagnum-dominated peatlands of North America. *J. Quat. Sci.* **23** 43–57.

BOOTH, R. K., LAMENTOWICZ, M. and CHARMAN, D. J. (2010). Preparation and analysis of testate amoebae in peatland paleoenvironmental studies. *Mires Peat* **7** 1–7.

BREWER, S., JACKSON, S. T. and WILLIAMS, J. W. (2012). Paleoecoinformatics: Applying geo-historical data to ecological questions. *Trends Ecol. Evol.* **27** 104–112.

CHARMAN, D. J. (2007). Summer water deficit variability controls on peatland water-table changes: Implications for Holocene palaeoclimate reconstructions. *Holocene* **17** 217–227.

CRESSIE, N. and JOHANNESSON, G. (2008). Fixed rank kriging for very large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 209–226. MR2412639

CSATÓ, L. (2002). Gaussian processes: Iterative sparse approximations. Ph.D. thesis, Aston Univ.

DATTA, A., BANERJEE, S., FINLEY, A. O. and GELFAND, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *J. Amer. Statist. Assoc.* **111** 800–812. MR3538706

DAWSON, A., PACIOREK, C. J., MCLACHLAN, J. S., GORING, S., WILLIAMS, J. W. and JACKSON, S. T. (2016). Quantifying pollen-vegetation relationships to reconstruct ancient forests using 19th-century forest composition and pollen data. *Quat. Sci. Rev.* **137** 156–175.

EDDELBUETTEL, D. and SANDERSON, C. (2014). RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Comput. Statist. Data Anal.* **71** 1054–1063. MR3132026

FINLEY, A. O., SANG, H., BANERJEE, S. and GELFAND, A. E. (2009). Improving the performance of predictive process modeling for large datasets. *Comput. Statist. Data Anal.* **53** 2873–2884. MR2667597

FUENTES, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *J. Amer. Statist. Assoc.* **102** 321–331. MR2345545

GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409. MR1141740

GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **1** 515–533. MR2221284

GELMAN, A., RUBIN, D. B. et al. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–472.

GNEITING, T. (2011). Making and evaluating point forecasts. *J. Amer. Statist. Assoc.* **106** 746–762. MR2847988

GNEITING, T., BALABDAOUI, F. and RAFTERY, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 243–268. MR2325275

GRANTHAM, N. S., REICH, B. J., PACIFICI, K., LABER, E. B., MENNINGER, H. L., HENLEY, J. B., BARBERÁN, A., LEFF, J. W., FIERER, N. et al. (2015). Fungi identify the geographic origin of dust samples. *PLoS ONE* **10** e0122605.

HASLETT, J., WHILEY, M., BHATTACHARYA, S., SALTER-TOWNSHEND, M., WILSON, S. P., ALLEN, J. R. M., HUNTLEY, B. and MITCHELL, F. J. G. (2006). Bayesian palaeoclimate reconstruction. *J. Roy. Statist. Soc. Ser. A* **169** 395–438. MR2236914

HEFLEY, T. J., BROMS, K. M., BROST, B. M., BUDERMAN, F. E., KAY, S. L., SCHARF, H. R., TIPTON, J. R., WILLIAMS, P. J. and HOOTEN, M. B. (2017). The basis function approach for modeling autocorrelation in ecological data. *Ecology* **98** 632–646.

HIGDON, D. (2002). Space and space-time modeling using process convolutions. In *Quantitative Methods for Current Environmental Issues* 37–56. Springer, London. MR2059819

JACKSON, S. T. and WILLIAMS, J. W. (2004). Modern analogs in Quaternary paleoecology: Here today, gone yesterday, gone tomorrow? *Annual Rev. Earth Planet. Sci.* **32** 495–537.

JUGGINS, S. (2015). rioja: Analysis of quaternary science data. R package version 0.9-9.

KAMMANN, E. E. and WAND, M. P. (2003). Geoadditive models. *J. Roy. Statist. Soc. Ser. C* **52** 1–18. MR1963210

LAUBER, C. L., HAMADY, M., KNIGHT, R. and FIERER, N. (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* **75** 5111–5120.

LEWANDOWSKI, D., KUROWICKA, D. and JOE, H. (2009). Generating random correlation matrices based on vines and extended onion method. *J. Multivariate Anal.* **100** 1989–2001. MR2543081

LINDGREN, F., RUE, H. and LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 423–498. MR2853727

MORALES-CASTILLA, I., MATIAS, M. G., GRAVEL, D. and ARAÚJO, M. B. (2015). Inferring biotic interactions from proxies. *Trends Ecol. Evol.* **30** 347–356.

MURRAY, I., ADAMS, R. P. and MACKAY, D. J. C. (2010). Elliptical slice sampling. In *AISTATS* **13** 541–548.

NOLAN, C., TIPTON, J., BOOTH, R. K., HOOTEN, M. B. and JACKSON, S. T. (2019). Comparing and improving methods for reconstructing peatland water-table depth from testate amoebae. *Holocene*.

NYCHKA, D., BANDYOPADHYAY, S., HAMMERLING, D., LINDGREN, F. and SAIN, S. (2015). A multiresolution Gaussian process model for the analysis of large spatial datasets. *J. Comput. Graph. Statist.* **24** 579–599. MR3357396

O'MALLEY, A. J. and ZASLAVSKY, A. M. (2008). Domain-level covariance analysis for multilevel survey data with structured nonresponse. *J. Amer. Statist. Assoc.* **103** 1405–1418. MR2655721

OVERPECK, J., WEBB, T. and PRENTICE, I. (1985). Quantitative interpretation of fossil pollen spectra: Dissimilarity coefficients and the method of modern analogs. *Quat. Res.* **23** 87–108.

OVERPECK, J. T., WEBB, R. S. and WEBB III, T. (1992). Mapping eastern North American vegetation change of the past 18 ka: No-analogs and the future. *Geology* **20** 1071–1074.

PACIOREK, C. J. (2007). Computational techniques for spatial logistic regression with large data sets. *Comput. Statist. Data Anal.* **51** 3631–3653. MR2364480

PACIOREK, C. J. and MCLACHLAN, J. S. (2009). Mapping ancient forests: Bayesian inference for spatio-temporal trends in forest composition using the fossil pollen proxy record. *J. Amer. Statist. Assoc.* **104** 608–622. MR2751442

PARNELL, A. C., SWEENEY, J., DOAN, T. K., SALTER-TOWNSHEND, M., ALLEN, J. R. M., HUNTLEY, B. and HASLETT, J. (2015). Bayesian inference for palaeoclimate with time uncertainty and stochastic volatility. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **64** 115–138. MR3293921

POURAHMADI, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* **86** 677–690. MR1723786

POURAHMADI, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika* **87** 425–435. MR1782488

PRISM CLIMATE GROUP, OREGON STATE UNIVERSITY. Available at http://prism.oregonstate.edu. Created February 4, 2004.

R CORE TEAM (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

ROBERTS, G. O. and ROSENTHAL, J. S. (2009). Examples of adaptive MCMC. *J. Comput. Graph. Statist.* **18** 349–367. MR2749836

SALTER-TOWNSHEND, M. and HASLETT, J. (2012). Fast inversion of a flexible regression model for multivariate pollen counts data. *Environmetrics* **23** 595–605. MR3020077

SAUCEDO-GARCÍA, A., ANAYA, A. L., ESPINOSA-GARCÍA, F. J. and GONZÁLEZ, M. C. (2014). Diversity and communities of foliar endophytic fungi from different agroecosystems of *Coffea arabica* L. in two regions of Veracruz, Mexico. *PLoS ONE* **9** e98454.

SCHAPIRE, R. E. (1990). The strength of weak learnability. *Mach. Learn.* **5** 197–227.

SEEGER, M. (2004). Low rank updates for the Cholesky decomposition. Technical Report No. EPFL-REPORT-161468.

SMITH, M. and KOHN, R. (2002). Parsimonious covariance matrix estimation for longitudinal data. *J. Amer. Statist. Assoc.* **97** 1141–1153. MR1951266

STAN DEVELOPMENT TEAM (2016). Stan Modeling Language Users Guide and Reference Manual, version 2.10.0 ed. Available at http://mc-stan.org.

STEIN, M. L. (1999). *Interpolation of Spatial Data*: *Some Theory for Kriging*. *Springer Series in Statistics*. Springer, New York. MR1697409

STEIN, M. L. (2014). Limitations on low rank approximations for covariance matrices of spatial data. *Spat. Stat.* **8** 1–19. MR3326818

TER BRAAK, C. J. F. and VAN DAM, H. (1989). Inferring pH from diatoms: A comparison of old and new calibration methods. *Hydrobiologia* **178** 209–223.

TIPTON, J. R, HOOTEN, M. B, NOLAN, C., BOOTH, R. K and MCLACHLAN, J. (2019). Supplement to "Predicting paleoclimate from compositional data using multivariate Gaussian process inverse prediction." DOI:10.1214/19-AOAS1281SUPP.

VASKO, K., TOIVONEN, H. T. and KORHOLA, A. (2000). A Bayesian multinomial Gaussian response model for organism-based environmental reconstruction. *J. Paleolimnol.* **24** 243–250.

VECCHIA, A. V. (1988). Estimation and model identification for continuous spatial processes. *J. Roy. Statist. Soc. Ser. B* **50** 297–312. MR0964183

WAHL, E. R., DIAZ, H. F. and OHLWEIN, C. (2012). A pollen-based reconstruction of summer temperature in central North America and implications for circulation patterns during medieval times. *Glob. Planet. Change* **84** 66–74.

WARTON, D. I., BLANCHET, F. G., O'HARA, R. B., OVASKAINEN, O., TASKINEN, S., WALKER, S. C. and HUI, F. K. C. (2015). So many variables: Joint modeling in community ecology. *Trends Ecol. Evol.* **30** 766–779.

WILLIAMS, J. and SHUMAN, B. (2008). Obtaining accurate and precise environmental reconstructions from the modern analog technique and North American surface pollen dataset. *Quat. Sci. Rev.* **27** 669–687.

J. R. TIPTON
DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF ARKANSAS
FAYETTEVILLE, ARKANSAS 72701
USA
E-MAIL: jrtipton@uark.edu

M. B. HOOTEN
U.S. GEOLOGICAL SURVEY
COLORADO COOPERATIVE FISH
AND WILDLIFE RESEARCH UNIT
DEPARTMENTS OF FISH, WILDLIFE
AND CONSERVATION BIOLOGY
AND
DEPARTMENT OF STATISTICS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80523
USA

C. NOLAN
DEPARTMENT OF GEOSCIENCES
UNIVERSITY OF ARIZONA
TUCSON, ARIZONA 85721
USA

R. K. BOOTH
EARTH AND ENVIRONMENTAL
SCIENCE DEPARTMENT
LEHIGH UNIVERSITY
BETHLEHEM, PENNSYLVANIA 18015
USA

J. MCLACHLAN
DEPARTMENT OF BIOLOGY
UNIVERSITY OF NOTRE DAME
NOTRE DAME, INDIANA 46556
USA