SPATIAL HYPERSPHERIC MODELS FOR COMPOSITIONAL DATA

BY MICHAEL R. SCHWOB[1,a] ![ORCID], MEVIN B. HOOTEN[2,b], NICHOLAS M. CALZADA[3,d] AND
TIMOTHY H. KEITT[4,c]

[1]*Department of Statistics, Virginia Tech,* [a]*schwob@vt.edu*

[2]*Department of Statistics and Data Sciences, The University of Texas at Austin,* [b]*mevin.hooten@austin.utexas.edu*

[3]*Department of Integrative Biology, The University of Texas at Austin,* [c]*tkeitt@utexas.edu*

[4]*Department of Statistics, The Ohio State University,* [d]*calzada.9@buckeyemail.osu*

Compositional observations are an increasingly prevalent data source in spatial statistics. Analysis of such data is typically done on log-ratio transformations or via Dirichlet regression. However, these approaches often make unnecessarily strong assumptions (e.g., strictly positive components, exclusively negative correlations). An alternative approach uses square-root transformed compositions and directional distributions. Such distributions naturally allow for zero-valued components and positive correlations, yet they may include support outside the nonnegative orthant and are not generative for compositional data. To overcome this challenge, we truncate the elliptically symmetric angular Gaussian (ESAG) distribution to the nonnegative orthant. Additionally, we propose a spatial hyperspheric regression model that contains fixed and random multivariate spatial effects. The proposed model also contains a term that can be used to propagate uncertainty that may arise from precursory stochastic models (i.e., machine learning classification). We used our model in a simulation study and for a spatial analysis of classified bioacoustic signals of the *Dryobates pubescens* (downy woodpecker).

1. Introduction. Compositional data are nonnegative and have a constant-sum constraint; they often take the form of percentages, probability vectors, or proportions (Aitchison (1982)). Such data have become prevalent in many scientific fields, such as genomics (Quinn et al. (2018), Gloor (2023)), microbiomics (Koslovsky (2023)), business (Greenacre et al. (2023), Pal and Heumann (2022)), demography (Martín-Fernández, Daunis-i-Estadella and Mateu-Figueras (2015)), forensics (Grantham et al. (2020)), ecology (Nolan et al. (2019), Stewart et al. (2022)), and environmental sciences (Tipton et al. (2019)). We are interested in analyzing bioacoustic classifications, which take the form of compositional data. A composition $\mathbf{u} = (u_1, \ldots, u_d)'$ can be represented as a point on the simplex

$$\boldsymbol{\Delta}^{d-1} = \left\{ (u_1, \ldots, u_d)' : u_k \geq 0 \; \forall k, \sum_{k=1}^{d} u_k = c \right\},$$

where c is the constant in the constant-sum constraint. Without loss of generality, we assume that the composition $\mathbf{u}$ has a unit-sum constraint and set $c = 1$. Most traditional statistical methods are only valid for data in Euclidean geometry. Therefore, special treatment is necessary when analyzing compositional data. Fortunately, the growth in compositional data sources has led to recent advances in compositional data analysis (CoDA).

Historically, compositional data were transformed to the Euclidean space through log-ratio transformations to make use of standard statistical techniques. However, this approach is problematic for compositions containing zeros (Aitchison (1982), Feng, Zhu and Steen-Adams (2015)). Despite this limitation, log-ratio transformations are still the most widely

used CoDA technique today (Greenacre (2022)). If compositional data contain zeros, the zeros are treated as missing and imputed or replaced with small quantities to facilitate log-ratio transformations; the underlying assumption in these cases is that the zeros are unobserved small quantities that were rounded to zero (Tsagris and Stewart (2018)). However, if the value is truly zero (a structural zero), such procedures are problematic because the modification of the zero-valued components implies a change in all components due to the unit-sum constraint. An alternative to the log-ratio transformation is the chiPower transformation, which allows for zero-valued components at the cost of slightly violating subcompositional coherence (Greenacre (2024)). Additionally, Leininger et al. (2013) proposed a spatial regression framework for compositional data that allows for zero-valued components while at least one component is nonzero for all observations; this is a weaker assumption but precludes some compositional data sets from analysis.

Another common method for CoDA is Dirichlet modeling. The Dirichlet distribution is a natural choice for CoDA because its support lies on the unit simplex. Early studies relying on Dirichlet distributions were challenged by compositions containing zeros, but recent work has overcome this challenge (Tsagris and Stewart (2018), Koslovsky (2023)). Notably, the Dirichlet distribution does not allow for positive correlations between components in a composition. The nested Dirichlet distribution, or the Dirichlet-tree distribution, is an extension that allows for positive correlations between components by constructing a tree of Dirichlet distributions in which components that fall under the same branch are positively correlated (Dennis (1991)). Typically, the structure of the tree is determined a priori or is chosen via model selection techniques; the former method assumes knowledge regarding which components should be positively correlated, and the latter method is computationally expensive for a moderately large d because many tree structures must be fit to the data (Rodríguez, Dunson and Gelfand (2008), Tang, Ma and Nicolae (2018)). Alternative approaches include the polynomially tilted pairwise interaction model (Scealy and Wood (2023)) and the power interaction model (Yu, Drton and Shojaie (2024)), which aim to infer pairwise interactions between components and are fit via score matching methods for the simplex. Notably, the Dirichlet distribution is a special case of the class of power interaction models.

A promising, yet understudied approach to CoDA is the transformation of compositional data to directional data. A composition $\mathbf{u}$ is transformed to a directional data point $\mathbf{y} = (y_1, \ldots, y_d)'$ with the square-root transformation $\mathbf{y} = \sqrt{\mathbf{u}}$, where the square-root function is applied elementwise. The resulting data point $\mathbf{y}$ exists on the surface of the unit hypersphere

$$\mathcal{S}^{d-1} = \left\{ \mathbf{y} \in \mathbb{R}^d : \|\mathbf{y}\|_2 = 1 \right\}$$

in the nonnegative orthant (see Figure 1), where $\| \cdot \|_2$ denotes the l_2-norm. After $\mathbf{y}$ is transformed, directional approaches may be used to analyze the data. Most distributions for directional data naturally allow for zeros and have flexible covariance structures that allow for positive and negative correlations between transformed components. Thus, directional distributions naturally overcome the challenges inherent in the log-ratio and Dirichlet-based approaches. One such distribution is the Kent distribution, or the 5-parameter Fisher–Bingham distribution. The Kent distribution is the most common distribution used to analyze directional data, despite having an analytically intractable normalizing constant for large d. Paine et al. (2018) recently proposed the elliptically symmetric angular Gaussian (ESAG) distribution, which has an analytically tractable normalizing constant. Then Paine et al. (2020) showed that the ESAG distribution provides inference similar to that obtained by the Kent distribution but at a fraction of the computational cost. We use the ESAG distribution to analyze transformed compositional data (i.e., bioacoustic classifications). Our approach naturally allows for zero-valued components and positive correlations between components, thereby accommodating all compositional data sets.
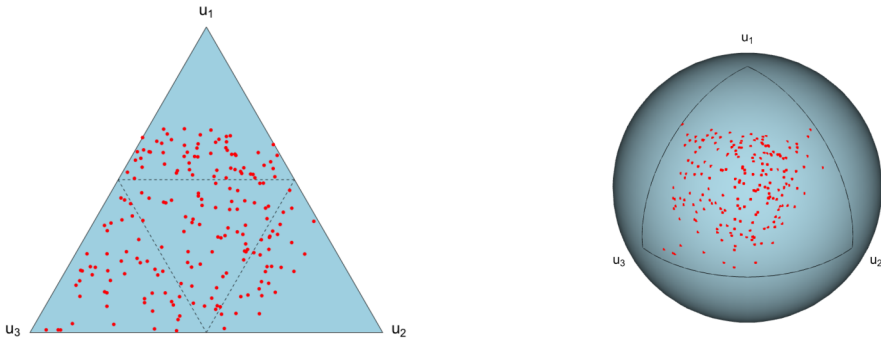
FIG. 1. *Left*: *A ternary plot of* $n = 200$ *compositions of* $d = 3$ *dimensions. Right*: *The square-root transformed compositions on the surface of a sphere in the nonnegative orthant* (*black lines*).

Analyzing the square-root transformed compositional data with conventional directional distributions is appropriate if the data are not distributed near the boundaries of the nonnegative orthant (Scealy and Welsh (2011)). However, if the transformed data are distributed near the boundaries, it is likely that directional distributions will place a nonnegligible amount of support outside the nonnegative orthant. Thus, directional distributions are not generative for compositional data because they may generate points outside the nonnegative orthant that cannot be transformed back to the simplex. In the directional statistics literature, there are generally two suggestions to remedy such boundary issues: (i) use the alternative transformation $|\mathbf{y}| = \sqrt{\mathbf{u}}$ and model $|\mathbf{y}|$ using a folded version of a directional distribution, or (ii) use a truncated directional distribution, which is ideal if the components of $\mathbf{u}$ contain a large number of zeros. The former remedy was implemented in Scealy and Welsh (2014), where the authors used a folded Kent distribution on the transformed data; they viewed the folding assumption as a missing data problem and inferred the missing sign resulting from the folding. Although the folding approach introduces potential bias in the form of sign imputation, it was the only feasible remedy historically; truncated directional distributions were computationally prohibitive due to intractable normalizing constants. However, the ESAG distribution can be truncated at a much smaller computational cost than most directional distributions. We propose the ESAG$^+$ distribution, where the ESAG is truncated to the nonnegative orthant. We then use the ESAG$^+$ distribution to analyze and predict bioacoustic classifications in a spatial setting.

One of the scientific fields generating the most compositional data is machine learning. In particular, machine learning classification is a new and readily adopted technique that results in compositional data in the form of classification (probability) vectors (Murphy (2022)). However, such classification output is rarely treated as a compositional data source for downstream statistical analyses. We analyze classification vectors and propagate the uncertainty from the machine learning classification in our downstream analysis of the compositional data.

In our downstream analysis, we use hyperspheric regression to analyze the transformed bioacoustic classifications. In general, hyperspheric regression links the mean direction of a directional distribution to a multivariate linear model. To our knowledge, we propose the first hyperspheric regression model with spatial covariates and latent spatial random effects. The spatial hyperspheric regression is expressed in a Bayesian hierarchical framework. With our proposed model, we conduct a spatial analysis of bioacoustic data comprising classifications of *Dryobates pubescens* (downy woodpecker) vocalizations.

The remainder of this paper is organized as follows. In Section 2, we describe the *D. pubescens* vocalizations as well as the machine learning methods that we use to classify the vocalizations to obtain compositional data. Then, we briefly review the ESAG distribution

and propose the ESAG$^+$ distribution in Section 3. In Section 4, we propose our spatial hyperspheric model to analyze bioacoustic classifications and provide a method for uncertainty propagation. We present the implementation details in Section 5 and a simulation study in Section 6. In Section 7, we return to the bioacoustic compositional analysis and discuss spatial compositional prediction. Finally, we present concluding remarks in Section 8.

2. Bioacoustic data. An increasingly popular source of compositional data are machine learning classification algorithms, which aim to classify an input based on its features. The output of the classification algorithm is a probability vector (i.e., a composition), where the ith element corresponds to the probability that the given input belongs to the ith class. We used machine learning to classify spectrograms of bioacoustic signals from the *Dryobates pubescens*. The resulting acoustic compositions were used in a downstream analysis to study the relationship between the environment and the acoustic behavior of *D. pubescens*.

Bioacoustic data are valuable for assessing biodiversity, habitat health, and animal behavior (Sueur and Farina (2015), Keitt and Abelson (2021)). In particular, avian bioacoustic signals are often used as bioindicators or phenological cues to study climate change (Farina and Gage (2017)). We analyze bioacoustic signals of the *D. pubescens* because they are non-migratory and do not learn new vocalization patterns from other birds (i.e., suboscine). These characteristics allow us to focus on the relationship between the environment and the acoustic behavior of the species. *D. pubescens* inhabit a large portion of the United States and Canada and are known to produce three distinct signals: a laugh-like whinnying call, drumming (i.e., a rapid, rhythmic pecking), and a high-pitched chirp known as a "pik" (Dodenhoff (2002)). Both sexes produce the three distinct signals. Whinnying plays a role in the formation of pair bonds and also serves as a territorial call (Kilham (1962), Mahan (1996)). Drumming is a conspecific signal used to establish and defend territories, attract mates, maintain contact with mates, and signal readiness for copulation (Mahan (1996)). Piking is a response to disturbance (Short and Sandström (1982)) and a contact vocalization associated with courtship flights (Kilham (1962)).

We gathered *D. pubescens* bioacoustic data from the Macaulay Library at the Cornell Lab of Ornithology, which is a publicly crowd-sourced repository of animal media (Betancourt and McLinn (2012)). For analysis, we retained the signals that were deemed "scientific quality" by the Macaulay Library, free of background noise, and had a three- to five-star rating by the Macaulay Library user community. Additionally, we focused on bioacoustic recordings from the Northeastern United States during the springs of 2020–2023. The resulting data set comprised 91 recordings, which are identified in Appendix A of the Supplementary Material (Schwob et al. (2025)).

We generated mel spectrograms from the recordings using the librosa Python package with a sample rate of 44.1 kHz (McFee et al. (2015)). A spectrogram is a visual depiction of the spectrum of frequencies of a signal over time, where the horizontal axis represents time, the vertical axis represents frequency, and the color intensity represents amplitude. The mel spectrogram is a spectrogram where the frequencies are converted to the mel scale, which is a perceptual scale of pitches designed to represent how humans perceive sound (Stevens, Volkmann and Newman (1937)). We converted the energy values of the mel spectrograms to a logarithmic decibel scale. Log-scaled mel spectrograms (LSMS) are often used in machine learning to capture information in a way that is more interpretable (Xu et al. (2018)). We used the LSMSs as image representations of the bioacoustic signals and as the input to the machine learning classification algorithm. Figure 2 shows three LSMSs for each signal type of the *D. pubescens*. We classified the 91 LSMSs with a two-dimensional convolutional neural network (2D-CNN) to obtain probability vectors for each recording. The 2D-CNN is commonly used to process avian acoustic signals (Xie and Zhu (2022)). Details on our 2D-CNN implementation are found in Appendix B (Schwob et al. (2025)). The resulting acoustic
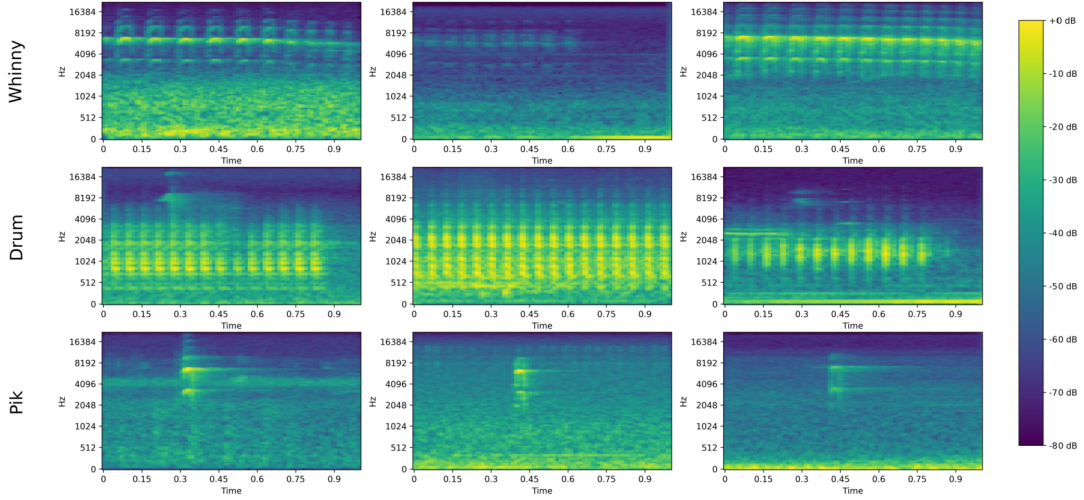
FIG. 2. *Three log-scaled mel spectrograms for each bioacoustic signal. These samples were randomly selected from the analyzed data.*

compositional data set did not contain zero-valued components, though other machine learning classifiers may render structural zeros. Finally, we squared the elements of the prediction vector to obtain the transformed directional data for downstream analysis.

3. ESAG distribution. We consider the $d \times 1$ random vector $\mathbf{y} = \mathbf{z}/\|\mathbf{z}\|_2$, where $\mathbf{z} \sim N_d(\boldsymbol{\mu}, \mathbf{V})$. The normalized random variable $\mathbf{y}$ follows an angular Gaussian distribution or a projected normal distribution, with support on $\mathcal{S}^{d-1}$. The angular Gaussian distribution is denoted as $AG(\boldsymbol{\mu}, \mathbf{V})$, where $\boldsymbol{\mu}$ is the mean direction and $\mathbf{V}$ controls the shape of the distribution on the surface of the hypersphere. The angular Gaussian parameters in $\boldsymbol{\mu}$ and $\mathbf{V}$ are not identifiable because $\mathbf{z}/\|\mathbf{z}\|_2$ and $c\mathbf{z}/\|c\mathbf{z}\|_2$ are equal for $c > 0$ and follow the same angular Gaussian distribution despite $\mathbf{z}$ and $c\mathbf{z}$ having different means and variances when $c \neq 1$. To impose identifiability, Paine et al. (2018) proposed the following sets of constraints: (i) $\mathbf{V}\boldsymbol{\mu} = \boldsymbol{\mu}$, and (ii) $\det(\mathbf{V}) = 1$, where $\det(\cdot)$ denotes the determinant of a matrix. These two sets of constraints lead to the construction of the ESAG distribution with probability density function

$$(1) \qquad [\mathbf{y}|\boldsymbol{\mu}, \mathbf{V}] = \frac{(2\pi)^{-(d-1)/2}}{(\mathbf{y}'\mathbf{V}^{-1}\mathbf{y})^{d/2}} \exp\left(\frac{1}{2}\left(\frac{(\mathbf{y}'\boldsymbol{\mu})^2}{\mathbf{y}'\mathbf{V}^{-1}\mathbf{y}} - \boldsymbol{\mu}'\boldsymbol{\mu}\right)\right) M_{d-1}\left\{\frac{\mathbf{y}'\boldsymbol{\mu}}{(\mathbf{y}'\mathbf{V}^{-1}\mathbf{y})^{1/2}}\right\},$$

where $M_{d-1}(t) = (2\pi)^{-1/2} \int_0^\infty x^{d-1} \exp\{-(x-t)^2/2\} \, dx$ and $[\cdot|\cdot]$ denotes the conditional probability density or mass function (Gelfand and Smith (1990)). We denote an ESAG-distributed random variable as $\mathbf{y} \sim ESAG_{d-1}(\boldsymbol{\mu}, \mathbf{V})$, where $\boldsymbol{\mu}$ is the mean direction and $\mathbf{V}$ determines the orientation of dispersion. Concentration is controlled by $\|\boldsymbol{\mu}\|_2$ with higher values leading to more concentration and less variability. Unlike the standard angular Gaussian distribution, the ESAG has elliptical symmetry that can capture anisotropic patterns in directional data. Additionally, the normalizing constant is much easier to compute than for many existing directional distributions. Finally, simulating data from angular Gaussian distributions is relatively fast, compared to most directional distributions, requiring only the normalization of a multivariate Gaussian simulation.

3.1. *An unconstrained parameterization.* The dimension of the parameter space for $N_d(\boldsymbol{\mu}, \mathbf{V})$ is $d(d+3)/2$, and the sets of constraints introduced by Paine et al. (2018) impose $d+1$ constraints. Thus, there are at most $\mathcal{P} = (d-1)(d+2)/2$ identifiable parameters for

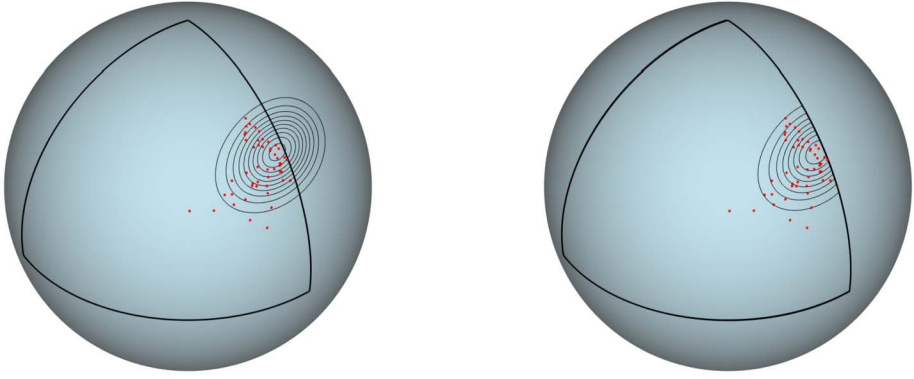FIG. 3. *A comparison of the support for an $ESAG_2$ (left) and $ESAG_2^+$ (right) distribution fit to transformed compositional data (dots) that are distributed near the boundary of the nonnegative orthant.*

$ESAG_{d-1}(\boldsymbol{\mu}, \mathbf{V})$. The original parameterization of the ESAG distribution imposed complex constraints on the model parameters. Yu and Huang (2024) introduced a new parameterization of the ESAG distribution that resulted in constraint-free model parameters in $\mathbb{R}$; we provide a brief review of their parameterization in Appendix C (Schwob et al. (2025)). In particular, $\mathbf{V}$ can be specified as a function of $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$, where $\boldsymbol{\gamma}$ are constraint-free real-valued vectors. We use this new parameterization in a Bayesian hierarchical model to simplify prior specification for unconstrained parameters. In the next section, we link the mean direction $\boldsymbol{\mu}$ to a hyperspheric regression model. For $\boldsymbol{\gamma}$, we assume the following prior for each element: $\gamma_{j,k} \sim \mathrm{N}(0, \sigma_\gamma^2)$ for $j = 1, \ldots, d-2$ and $k = 1, \ldots, j+1$.

3.2. *The truncated ESAG distribution.* When compositional data are transformed to directional data under the square-root transformation, they lie on the surface of the nonnegative orthant of the hypersphere. Conventional directional distributions may generate points outside of the nonnegative orthant, which cannot be transformed back to the simplex via an elementwise squaring. Therefore, when analyzing transformed compositional data (i.e., bioacoustic classification vectors), we truncate the ESAG distribution to the nonnegative orthant, denoted $ESAG_{d-1}^+$, which has probability density function

$$(2) \qquad [\mathbf{y}|\boldsymbol{\mu}, \mathbf{V}]^+ = \frac{[\mathbf{y}|\boldsymbol{\mu}, \mathbf{V}]\mathbb{1}(\mathbf{y} \in \mathcal{S}_+^{d-1})}{\int_{\mathcal{S}_+^{d-1}} [\mathbf{y}|\boldsymbol{\mu}, \mathbf{V}] \, d\mathbf{y}},$$

where $[\mathbf{y}|\boldsymbol{\mu}, \mathbf{V}]$ is the $ESAG_{d-1}$ density function from (1), $\mathcal{S}_+^{d-1}$ denotes the surface of the nonnegative orthant of the unit hypersphere of d dimensions and $\mathbb{1}(\cdot)$ is an indicator function that evaluates to 1 if the condition is satisfied and 0 otherwise. Due to the truncation, $\boldsymbol{\mu}$ is no longer the mean direction, although it is still a location parameter; see Figure 3 for an example of an $ESAG_2$ distribution with support outside of the nonnegative orthant and our corresponding $ESAG_2^+$ distribution.

For many directional distributions, evaluation of the normalizing constant is computationally prohibitive. However, the normalizing constant associated with the ESAG distribution can be quickly approximated using Monte Carlo integration because simulating from the ESAG is fast. Thus, we approximate the normalizing constant for the $ESAG_{d-1}^+$ distribution with

$$(3) \qquad \int_{\mathcal{S}_+^{d-1}} [\mathbf{y}|\boldsymbol{\mu}, \mathbf{V}] \, d\mathbf{y} \approx \frac{1}{M} \sum_{i=1}^{M} \mathbb{1}(\mathbf{y}_i^* \in \mathcal{S}_+^{d-1}),$$

where $\mathbf{y}_i^*$ is a random draw from $\text{ESAG}_{d-1}(\boldsymbol{\mu}, \mathbf{V})$ for $i = 1, \ldots, M$. A quick method to simulate $\mathbf{y}_i^*$ is to simulate and normalize $\mathbf{z}_i^* \sim \text{N}_d(\boldsymbol{\mu}, \mathbf{V})$. For moderately small d, we find that $M = 5000$ Monte Carlo draws are sufficient to adequately approximate the normalizing constant. We explore the accuracy and computational efficiency of the approximation in (3) for various d and M in Appendix D (Schwob et al. (2025)).

Under the ESAG^+ distribution, the expected response $\text{E}(\mathbf{y}) = \mathbf{y}^*$ has two sets of constraints: (i) $\|\mathbf{y}^*\|_2 = 1$, and (ii) $0 \leq y_i^* \leq 1$ for $i = 1, \ldots, d$. These constraints are similar to the constraints imposed by a Dirichlet regression parameterized via the mean $\text{E}(\mathbf{u})$, where $\mathbf{u}$ is the compositional response: $\sum_{i=1}^d \text{E}(u_i) = 1$, and $0 \leq \text{E}(u_i) \leq 1$ for $i = 1, \ldots, d$. Under such constraints, regression parameters may not be identifiable because knowledge of $d - 1$ components implies knowledge of the dth component. To induce identifiability, one component is often taken as the baseline component, and its regression coefficients are fixed at zero (Douma and Weedon (2019)). Baseline components (i.e., reference categories) have been used for compositional data analysis with log-ratio transformations (Greenacre et al. (2023)), power scaling models (Leininger et al. (2013)), polynomially tilted pairwise interaction models (Scealy and Wood (2023)), and Dirichlet regression (Douma and Weedon (2019)). The constraints induced by the ESAG^+ distribution imply that the dth component of $\mathbf{y}^*$ can be obtained given the other $d - 1$ components. Thus, we establish a baseline component when using truncated directional distributions for compositional data analysis to induce identifiability for the regression parameters. Common choices for the baseline component are the first component or the most prevalent component in the data. For our bioacoustic compositional analysis, we selected the baseline category as "drumming" because it was the most commonly predicted *D. pubescens* signal in the classification algorithm.

4. Hyperspheric regression.

Hyperspheric regression is becoming a popular tool to analyze compositional data (Scealy and Welsh (2011), Yu and Huang (2025)) and directional data (Wang and Gelfand (2013), Scealy and Wood (2019), Paine et al. (2020), Garcia-Fernandez et al. (2021)). The general aim is to estimate the relationship between covariates of interest and the mean direction of directional distributions; this is analogous to conventional multivariate regression approaches in Euclidean space, yet hyperspheric regression accounts for the d-dimensional spherical geometry in which directional data and square-root transformed compositional data reside. Recently, Yu and Huang (2025) proposed the first spherical regression framework featuring an ESAG data model. In this section, we extend their work by considering a Bayesian hierarchical hyperspheric regression model that accounts for fixed and random spatial effects. We also account for uncertainty propagation because our bioacoustic classifications arise as output from a precursory stochastic model (i.e., machine learning classification).

4.1. *Spatial hyperspheric regression.* We specify a Bayesian hierarchical model that regresses directional data (i.e., transformed bioacoustic compositions) on spatial covariates and latent spatial random effects. We let $\mathbf{y}_i \sim \text{ESAG}_{d-1}^+(\boldsymbol{\mu}(\mathbf{x}_i), \mathbf{V}(\mathbf{x}_i))$ for $i = 1, \ldots, n$, where n and d denote the number of observations and the dimension of each observation, respectively. As in most first-moment regression frameworks, the location parameter $\boldsymbol{\mu}$ is specified as a function of the covariates $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p})'$. Because $\mathbf{V}$ is a function of $\boldsymbol{\mu}$ in the new parameterization, we explicitly note that $\mathbf{V}(\mathbf{x}_i)$ is also a function of the spatial covariates, so the orientation of dispersion is spatially heterogeneous.

In the standard ESAG distribution, $\boldsymbol{\mu}(\mathbf{x}_i)$ is the mean direction of the data, which lies in the nonnegative orthant if the directional data are transformed compositional data. Therefore, we use the softplus link function to ensure that all elements of $\boldsymbol{\mu}(\mathbf{x}_i)$ are nonnegative,

$$(4) \qquad \boldsymbol{\mu}(\mathbf{x}_i) = \text{softplus}(\mathbf{h}(\mathbf{x}_i)),$$

where softplus$(x) = \log(1 + e^x)$ is applied elementwise and $\mathbf{h}(\mathbf{x}_i)$ takes the form of a multivariate linear regression model. We use the softplus link function, instead of the more conventional log-link function, because we found that, as the range of concentration (i.e, $\{\|\boldsymbol{\mu}(\mathbf{x}_i)\|_2\}_{i=1}^n$) expands, mixing gets worse. In particular, the quasi-additive nature of the softplus function generally leads to better mixing than the multiplicative nature of the log-link function (Wiemann, Kneib and Hambuckers (2024)). Additionally, the softplus link function allows for a near additive interpretation of the relationship between the covariates and $\boldsymbol{\mu}(\mathbf{x}_i)$, which is generally more intuitive.

For our spatial hyperspheric regression model, we specify

$$(5) \qquad \boldsymbol{\mu}(\mathbf{x}_i) = \text{softplus}(\mathbf{B}\mathbf{x}_i + \boldsymbol{\eta}(\mathbf{s}_i)),$$

where $\mathbf{B}$ is a $d \times p$ matrix of regression coefficients and $\boldsymbol{\eta}(\mathbf{s}_i)$ are the $d \times 1$ latent spatial random effects at the location of the ith observation, $\mathbf{s}_i$. We denote the jth row of $\mathbf{B}$ as $\mathbf{B}_{j,\cdot} = \boldsymbol{\beta}_j' = (\beta_{j,1}, \ldots, \beta_{j,p})$ and assume a multivariate Gaussian prior for each row: $\boldsymbol{\beta}_j \sim$ $\text{N}(\mathbf{0}_p, \sigma_\beta^2 \mathbf{I}_p)$. The latent effects $\boldsymbol{\eta}(\mathbf{s}_i)$ suggest that each component (i.e., bioacoustic signal) has its own spatial field.

Perhaps the simplest approach to modeling $\boldsymbol{\eta}(\mathbf{s}_i)$ is to assume that the latent random effects comprise d independent spatial fields. However, we assume that the spatial fields among the components are dependent and model $\boldsymbol{\eta}(\mathbf{s}_i)$ with a multivariate spatial process. Ideally, one would have knowledge about how one component may depend on another; if so, a conditional multivariate spatial process may be specified (Cressie and Wikle (2011)). We assume a lack of such knowledge and specify $\boldsymbol{\eta}(\mathbf{s}_i)$ jointly. We model the multivariate spatial process with a linear model of coregionalization (LMC), which provides a flexible framework that can capture complex dependence structures (Schmidt and Gelfand (2003), Gelfand and Banerjee (2010)). In particular, we specify the matrix-variate LMC prior $\mathbf{H} \sim \text{LMC}(\mathbf{C}, \mathbf{R}_{j=1}^d)$ with density function

$$(6) \qquad p(\mathbf{H}|\mathbf{C}, \mathbf{R}_{j=1}^d) = \frac{\exp(-\frac{1}{2}\sum_{j=1}^d \mathbf{c}_j^{-1}\mathbf{H}\mathbf{R}_j^{-1}\mathbf{H}'(\mathbf{c}_j^{-1})')}{(2\pi)^{nd/2}|\det(\mathbf{C})|^n \prod_{j=1}^d \det(\mathbf{R}_j)^{1/2}},$$

where $\mathbf{H}$ is a $d \times n$ matrix with columns $\boldsymbol{\eta}(\mathbf{s}_1), \ldots, \boldsymbol{\eta}(\mathbf{s}_n)$, the $\{\mathbf{R}_j\}$ are spatial correlation matrices obtained by applying the correlation functions $\rho(|\mathbf{s}_i - \mathbf{s}_k|; \phi_j)$ to all pairs of locations, $\mathbf{c}_j$ is the jth column of the coregionalization coefficients matrix $\mathbf{C}$, and $\mathbf{c}_j^{-1}$ is the jth row of $\mathbf{C}^{-1}$. If observations are co-located, there are $m < n$ distinct locations; without loss of generality, we let $m = n$. Alie, Stephens and Schmidt (2024) show that (6) is equivalent to

$$\text{vec}(\mathbf{H}) = \boldsymbol{\eta} \sim \text{N}_{dn}\left(\mathbf{0}, \sum_{j=1}^d \mathbf{R}_j \otimes \mathbf{c}_j\mathbf{c}_j'\right),$$

where

$$(7) \qquad \boldsymbol{\eta} = \big(\eta(\mathbf{s}_1)_1, \ldots, \eta(\mathbf{s}_1)_d, \ldots, \eta(\mathbf{s}_n)_1, \ldots, \eta(\mathbf{s}_n)_d\big)'$$

is the $dn \times 1$ stacked vector of latent spatial random effects and $\otimes$ denotes the Kronecker product. We use the matrix-variate parameterization of the LMC due to its computational savings (Zhang and Banerjee (2022)). In particular, computation can be executed directly on $\mathbf{C}$ and $\{\mathbf{R}_j\}$ rather than computing the determinant and inverse of the $nd \times nd$ quadratic product under the vectorized form (Kyzyurova (2017)).

The coregionalization matrix $\mathbf{C}\mathbf{C}'$ captures the relationships between spatially correlated variables and decomposes the overall variance into contributions from individual variables and their covariances (Schmidt and Gelfand (2003)). A common approach to modeling $\mathbf{C}$

is to specify a lower triangular parameterization to induce identifiability. However, such a constraint imposes an asymmetry on the components of the model, which is limiting and often unjustifiable (Alie, Stephens and Schmidt (2024)). Although $\mathbf{C}$ is not identifiable, the positive definite matrix $\mathbf{CC}'$ is identifiable, which is sufficient for the identifiability of the multivariate spatial process $\boldsymbol{\eta}$. We assume a lack of knowledge about across-component covariance and specify independent standard normal priors for each element of $\mathbf{C}$.

We construct the $n \times n$ spatial correlation matrices by specifying the Matérn 3/2 correlation function

$$\rho\big(|\mathbf{s}_i - \mathbf{s}_k|; \phi_j\big) = \left(1 + \frac{\sqrt{3}|\mathbf{s}_i - \mathbf{s}_k|}{\phi_j}\right) \exp\left(-\frac{\sqrt{3}|\mathbf{s}_i - \mathbf{s}_k|}{\phi_j}\right)$$

for $j = 1, \ldots, d$, where $|\mathbf{s}_i - \mathbf{s}_k|$ is the Euclidean distance between locations $\mathbf{s}_i$ and $\mathbf{s}_k$. We model spatial dependency with the Matérn 3/2 correlation function because it offers flexibility in modeling smoothness, providing a balance between smooth functions (such as Gaussian decay) and rougher functions (such as exponential decay). Additionally, the moderate complexity of the Matérn 3/2 function reduces the risk of overfitting compared to higher-order Matérn correlation functions (Hoeting et al. (2006)). We value the flexibility in modeling smoothness because we assume a lack of knowledge about the smoothness of bioacoustic signal classification. In other applications the selection of $\rho(|\mathbf{s}_i - \mathbf{s}_k|; \phi)$ can be aided by domain expertise, exploratory data analysis, and model selection. For our bioacoustic analysis, we assign independent informative priors $\phi_j \sim \text{Gamma}(62, 0.04)$, which correspond to practical ranges at distances between 0.01 and half of the maximum observed distance in our bioacoustic data. We use informative priors because the spatial range parameters cannot be consistently estimated along with the marginal variances contained in $\mathbf{C}$ (Zhang (2004)).

4.2. *Uncertainty propagation.* We consider the case where the compositional or directional data are obtained as output from another model; in our bioacoustic application, the compositions are obtained from a machine learning classifier, which inherently introduces uncertainty. When output from model A is treated as input to model B, it is typically desirable to account for any uncertainty associated with the output of model A in model B (Draper (1995)). We account for such uncertainty by introducing an additive term to the hyperspheric regression model,

$$(8) \qquad \boldsymbol{\mu}(\mathbf{x}_i, \mathbf{z}_i) = \text{softplus}\big(\mathbf{B}\mathbf{x}_i + \boldsymbol{\eta}(\mathbf{s}_i) + \mathbf{A}\mathbf{z}_i\big),$$

where $\mathbf{z}_i$ is a q-dimensional vector of covariates used to quantify uncertainty in model A predictions, $\mathbf{A} = \mathbf{1}_d \otimes \boldsymbol{\alpha}'$ is a $d \times q$ matrix, $\mathbf{1}_d$ is a d-dimensional vector of ones, and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)'$ are regression coefficients for $\{\mathbf{z}_i\}_{i=1}^n$. The role of $\mathbf{A}\mathbf{z}_i$ is to attenuate the concentration of the data model at the observation level. If an observation was predicted by model A with a high level of uncertainty, then $\mathbf{A}\mathbf{z}_i$ is expected to be smaller for that observation (i.e., decreasing the concentration). Similarly, if an observation was predicted by model A with a high level of certainty, then $\mathbf{A}\mathbf{z}_i$ is expected to be larger (i.e., increasing the concentration). The uncertainty propagation term $\mathbf{A}\mathbf{z}_i$ must be included in the specification for $\boldsymbol{\mu}$ because its l_2-norm specifies the concentration. However, we do not want $\mathbf{A}\mathbf{z}_i$ to contribute to the learning of the normalized mean direction (i.e., to affect the point at which the mean direction intersects the surface of the nonnegative orthant). Thus, we specify $\mathbf{A}$ to have identical rows comprising $\boldsymbol{\alpha}$. We consider the prior $\boldsymbol{\alpha} \sim \text{N}(\mathbf{0}_q, \sigma_\alpha^2 \mathbf{I}_q)$ in our Bayesian hierarchical model. The complete model specification is outlined in Appendix E (Schwob et al. (2025)).

5. Implementation. We fit the proposed Bayesian hierarchical hyperspheric regression model using an MCMC algorithm. We used Metropolis–Hastings updates for $\boldsymbol{\alpha}$, $\{\boldsymbol{\beta}_j\}_{j=1}^d$, $\boldsymbol{\gamma}$, $\{\phi_j\}_{j=1}^d$, and the elements of $\mathbf{C}$. For $\boldsymbol{\alpha}$, we used a multivariate symmetric random-walk proposal: $\boldsymbol{\alpha}^{(*)} \sim \mathrm{N}(\boldsymbol{\alpha}^{(k-1)}, \sigma_\alpha^2 \mathbf{I})$, where $(*)$ denotes the proposed values and $(k-1)$ denotes the values on the previous MCMC iteration. We also used a symmetric random-walk proposal for each row of $\mathbf{B}$, each element of $\boldsymbol{\gamma}$, and each element of $\mathbf{C}$,

$$\boldsymbol{\beta}_j^{(*)} \sim \mathrm{N}(\boldsymbol{\beta}_j^{(k-1)}, \sigma_\beta^2 \mathbf{I}), \quad j = 1, \ldots, d,$$

$$\gamma_{l,i}^{(*)} \sim \mathrm{N}(\gamma_{l,i}^{(k-1)}, \sigma_\gamma^2), \quad l = 1, \ldots, d-2, i = 1, \ldots, l+1,$$

$$C_{ij}^{(*)} \sim \mathrm{N}(C_{ij}^{(k-1)}, \sigma_c^2), \quad i = 1, \ldots, d, j = 1, \ldots, d,$$

where we only accepted the proposal $C_{ij}^{(*)}$ if the resulting coregionalization matrix $\mathbf{CC}'$ was positive definite (Gelfand, Smith and Lee (1992)). We used the proposal $\phi_j^{(*)} \sim \mathrm{N}(\phi_j^{(k-1)}, \sigma_\phi^2)_0^\infty$ with a random-walk truncated below 0 to ensure the spatial range was non-negative. Finally, we updated the nd-dimensional vector $\boldsymbol{\eta}$ with an elliptical slice sampler, which provided significantly better mixing than Metropolis algorithms for moderately high dimensions (Murray, Adams and MacKay (2010)). Then we constructed $\mathbf{H}$ from $\boldsymbol{\eta}$, which was used to evaluate the matrix-variate LMC density in (6) during the updates for $\{\phi_j\}_{j=1}^d$ and the elements of $\mathbf{C}$.

For computational efficiency, we implemented the MCMC algorithm in Julia and used static arrays to significantly decrease computation costs (Bezanson et al. (2017)). We also evaluated the $M_{d-1}(\cdot)$ term of the ESAG and ESAG$^+$ densities using adaptive Gauss–Kronrod quadrature, which balanced computational speed and accuracy (Johnson (2013)). Finally, we approximated the normalizing constant for the ESAG$^+$ density in (3) using parallelized Monte Carlo integration. In both the simulation study and the bioacoustic analysis, we ran the MCMC algorithm until the $\hat{R}$ convergence diagnostic was less than 1.01 for all parameters in the model; the spatial range parameters $\{\phi_j\}$ required the most iterations to satisfy this condition (Zhang (2004)). We ran our MCMC algorithm for 300,000 iterations and 400,000 iterations for the simulation study and bioacoustic analysis, respectively. The computational costs of our model are provided for the simulation study and the bioacoustic application in the respective sections.

5.1. *Compositional inference.* If the original data are compositional (i.e., bioacoustic classifications), there may be interest in predicting a new composition, given a set of covariates. Yu and Huang (2024) suggested that predicting a composition is equivalent to estimating $\mathrm{E}(\mathbf{y}^2) = \boldsymbol{\Xi}^2 \mathrm{E}(\mathbf{K}^2)$, where $\boldsymbol{\Xi} = [\boldsymbol{\xi}_d \, \boldsymbol{\xi}_{d-1} \ldots \boldsymbol{\xi}_1]$, $\mathbf{K} = \boldsymbol{\Xi}' \mathbf{y}$, and the square is applied element-wise. The $\{\boldsymbol{\xi}\}_{i=1}^d$ are defined in Appendix C as the orthonormal eigenvectors of $\mathbf{V}$ and can be obtained deterministically, given $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ (Schwob et al. (2025)). A Monte Carlo estimator for $\mathrm{E}(\mathbf{y}^2)$ is

$$\widehat{\mathrm{E}(\mathbf{y}^2)} = \widehat{\boldsymbol{\Xi}}^2 \sum_{i=1}^M \widehat{\mathbf{K}}_i^2 / M,$$

where M is the number of Monte Carlo simulations, $\widehat{\mathbf{K}}_i = \widehat{\boldsymbol{\Xi}}' \mathbf{y}_i$, and $\widehat{\boldsymbol{\Xi}}$ is the posterior mean of $\boldsymbol{\Xi}$. An alternative estimator approximates $\mathrm{E}(\mathbf{y}^2)$ via Monte Carlo integration,

$$(9) \qquad \int_{\mathcal{S}_{d-1}^+} \mathbf{y}^2 \cdot [\mathbf{y}|\boldsymbol{\mu}, \mathbf{V}]^+ \, d\mathbf{y} \approx \frac{1}{M} \sum_{i=1}^M \mathbf{y}_i^2 \cdot [\mathbf{y}_i|\hat{\boldsymbol{\mu}}, \hat{\mathbf{V}}]^+,$$

where $[\mathbf{y}_i | \boldsymbol{\mu}, \mathbf{V}]^+$ is the probability density function for the ESAG$^+$ distribution in (2), $\hat{\boldsymbol{\mu}}$ is the posterior mean of $\boldsymbol{\mu}$, and $\hat{\mathbf{V}} = f(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\gamma}})$, where $\hat{\boldsymbol{\gamma}}$ is the posterior mean of $\boldsymbol{\gamma}$ and $f(\cdot)$ is specified in Appendix C (Schwob et al. (2025)). Uniformly random samples from $\mathcal{S}_{d-1}^+$ can be obtained by computing $\mathbf{y}_i = |\mathbf{z}_i| / \|\mathbf{z}_i\|_2$, where $|\cdot|$ is applied elementwise and $\mathbf{z}_i \sim \mathrm{N}(\mathbf{0}_d, \mathbf{I}_d)$ for $i = 1, \ldots, M$.

6. Simulation study. We simulated directional data from our Bayesian hierarchical model and aimed to recover the true parameter values. First, we uniformly sampled $n = 100$ observations from the study domain $\mathcal{S} = [0, 1] \times [0, 1]$. The covariates for each observation were fully specified by their spatial location: $\mathbf{x}(\mathbf{s}_i) = (1, x(\mathbf{s}_i)_1, x(\mathbf{s}_i)_2)'$ for $i = 1, \ldots, 100$, where $x(\mathbf{s}_i)_1 = |s_{i,1} - 0.5|^{1.2}$ and $x(\mathbf{s}_i)_2 = \|\mathbf{s}_i\|_2$. Then we standardized the environmental covariates to be in $[1, 2]$, which is more comparable in scale with directional data (Scealy and Wood (2019)). We set the first of three dimensions as the baseline component and specified

$$\mathbf{B} = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 1.25 & 1.15 & 0.25 \\ 1.1 & 0.55 & 1.65 \end{bmatrix}, \qquad \mathbf{C} = \begin{bmatrix} 1.46 & -0.91 & 0.43 \\ 1.15 & -0.60 & 1.35 \\ 1.18 & 1.54 & 1.09 \end{bmatrix},$$

$\boldsymbol{\phi} = (0.12, 0.15, 0.19)'$, $\alpha = -0.31$, and $\boldsymbol{\gamma} = (0.728, 0.346)'$. Then we simulated $\boldsymbol{\eta}$, computed $\boldsymbol{\mu}(\mathbf{x}_i)$ and $\mathbf{V}(\mathbf{x}_i)$, and simulated the directional response $\mathbf{y}_i$ from the ESAG$^+$ data model. We specified the following hyperparameters: $\sigma_\alpha^2 = 10$, $\sigma_\beta^2 = 10$, and $\sigma_\gamma^2 = 10$. We specified the hyperparameters in $\phi_j \sim \mathrm{Gamma}(\alpha_\phi, \theta_\phi)$ for $j = 1, \ldots, d$ such that 95% of the prior support was less than half of the observed maximum distance with a standard deviation of 0.03.

We fit two versions of our Bayesian hierarchical model to the simulated data: (i) with an ESAG$^+$ data model with the first component set as the baseline and (ii) with an ESAG data model (which did not need a baseline component). The models were fit to the simulated data with an MCMC algorithm that ran for 300,000 iterations with a burn-in of 60,000 iterations. The ESAG model and ESAG$^+$ model performed 41 it/s and 6 it/s, respectively, on a 3.2 GHz processor with 64 GB of RAM. For the ESAG$^+$ model, the effective sample sizes for $\{\phi_j\}$ were 1725.213, 1728.566, and 1694.102. The average effective sample size for $\boldsymbol{\eta}$ was 1693.296 with a minimum of 1620.348 and a maximum of 1735.055. The effective sample sizes under the ESAG model were comparable.

Both models captured the true values for α and $\boldsymbol{\gamma}$ in their 95% credible intervals. Only the ESAG$^+$ model captured $\boldsymbol{\phi}$ in 95% credible intervals. However, the spatial range parameters are challenging to estimate for multivariate spatial processes (Gelfand et al. (2004), Zhang (2004)). The true regression coefficients $\mathbf{B}$ were recovered by both models, though the posterior means were closest under the ESAG$^+$ model. For both models we estimated the normalized mean direction $\mathrm{E}(\mathbf{y})$ across a fine grid spanning the study domain. The posterior normalized mean directions are compared to the true normalized mean directions in Figure 4, which visually indicates that the ESAG$^+$ model better recovered the truth as expected.

For quantitative model comparison, we used the logarithmic score function (logS) and the χ^2 measure of compositional distance (CSD). The logS is commonly used to score multivariate predictions because it evaluates the mean and variance of the marginal distribution and the correlation structure of the prediction (Bjerregård, Møller and Madsen (2021)). The mean logS were 3.265 and 1.621 for the ESAG and ESAG$^+$ models, respectively, indicating that the ESAG$^+$ model provided a better fit to the data. Out of a set of compositional distance measures that were scale and permutation invariant and allowed for zero-valued components, Stewart (2017) found that the CSD was the measure that violated subcompositional coherence the least. We used CSD to compare the squared observations with the predicted compositions of the competing models. In particular, we used (9) to predict the composition for the ESAG$^+$
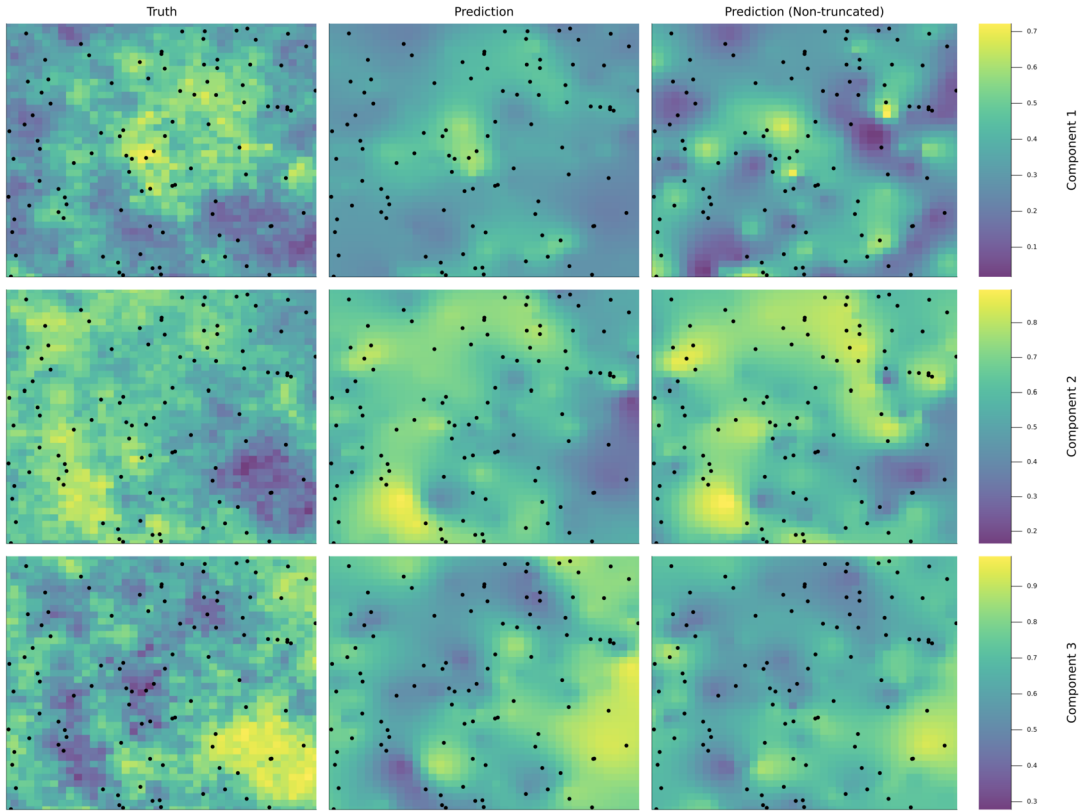
FIG. 4. *The true normalized mean direction (left column) and the posterior normalized mean directions from the ESAG$^+$ model (center column) and ESAG model (right column). The dots denote the observed locations.*

model and a similar Monte Carlo integration containing the nontruncated density to predict the composition for the ESAG model. The mean CSDs were 0.603 and 0.542 for the ESAG and ESAG$^+$ models, respectively, further suggesting that the ESAG$^+$ model provided the better fit to the data.

7. Bioacoustic compositional analysis. We fit our Bayesian hierarchical model in Appendix E to the *D. pubescens* bioacoustic data (Schwob et al. (2025)). We considered two environmental covariates that we hypothesized influence *D. pubescens* acoustic behavior: temperature and precipitation (Dodenhoff (2002)). In particular, we collected the average temperature in degrees Celsius ($\mathbf{x}_1$) and precipitation in millimeters ($\mathbf{x}_2$) at the observed locations across the spring (March–May) in which the bioacoustic signals were recorded. Similar to Scealy and Wood (2019), we transformed the covariates with $x_{i,k}^* = (x_{i,k} - x_{(1),k})/(x_{(n),k} - x_{(1),k}) + 1$ for $i = 1, \ldots, n$, where $x_{(1),k}$ and $x_{(n),k}$ are the minimum and maximum ordered statistics for the kth covariate, respectively. The transformed covariates were more comparable in scale with the response of a unit Euclidean norm. The final design matrix was $\mathbf{X} = [\mathbf{1}_n \, \mathbf{x}_1^* \, \mathbf{x}_2^*]$.

For uncertainty propagation, we considered the Macaulay Library user community ratings to be the auxiliary information $\mathbf{z}_i = (z_{i,1}, z_{i,2})'$, where $z_{i,1}$ and $z_{i,2}$ are indicator covariates that equal 1 if the community rating is four stars or three stars, respectively. We used community ratings as the auxiliary information because lower quality data were assumed to have more noise and lead to less confident (more variable) predictions in the machine learning classification.

We specified the following hyperparameters in our model: $\sigma_\alpha^2 = 10$, $\sigma_\beta^2 = 10$, and $\sigma_\gamma^2 = 10$. Such variances are vague for an ESAG regression because the magnitude of the regression parameters contribute quadratically in d to the l_2-norm of $\boldsymbol{\mu}$, which is typically small. For the spatial range parameters, we assigned independent informative priors $\phi_j \sim \text{Gamma}(62, 0.04)$, which corresponded to practical ranges of avian soundscapes at distances between 0.01 and half of the maximum observed distance in the data (Farina and Gage (2017)). Finally, we specified random-walk proposals with variances of 0.1, 0.1, 0.1, 0.01, and 0.5 for $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, $\mathbf{C}$, $\boldsymbol{\phi}$, and $\boldsymbol{\beta}$, respectively. We implemented the hierarchical model in an MCMC algorithm in Julia, which we ran for 400,000 iterations with a burn-in of 160,000 iterations. We used $M = 5000$ Monte Carlo draws for approximating the normalizing constant in (3). The MCMC algorithm performed 8 it/s on a 3.2 GHz processor with 64 GB of RAM. The effective sample sizes for $\{\phi_j\}$ were 1515.658, 1506.136, and 1508.484. The average effective sample size for $\boldsymbol{\eta}$ was 2321.221 with a minimum of 1571.198 and a maximum of 3169.788.

The posterior mean for $\boldsymbol{\alpha}$ was $\widehat{\boldsymbol{\alpha}} = (-0.62516, -1.02445)'$. Thus, lower community ratings were estimated to decrease the concentration of $\boldsymbol{\mu}$, thereby propagating more uncertainty in the data model; this was expected because it was assumed that community ratings were negatively correlated with noise. The posterior mean for $\mathbf{B}$ was

$$\widehat{\mathbf{B}} = \begin{bmatrix} -1.05524 & -1.11464 & -0.370964 \\ -5.17297 & 0.0 & 0.0 \\ 0.847705 & -0.396902 & -2.22877 \end{bmatrix},$$

where the rows correspond to the environmental effects of identifying whinnying, drumming, and piking, respectively. Inference on the environmental effects must be made relative to the baseline category (i.e., drumming). The magnitude of the intercept for the drumming category was quite large to offset the exclusively negative relative relationships that the other signals have with the environmental covariates. We did not draw inference from $\widehat{\mathbf{B}}$ directly because $\boldsymbol{\mu}$ is not the mean direction under the ESAG$^+$ model. Rather, we predicted classifications of *D. pubescens* bioacoustic signals across the study domain and compared the predictions to maps of the environmental covariates.

7.1. *Spatial compositional prediction.* We predicted the bioacoustic classification compositions across the Northeastern United States for each spring during the study period. To do this, we collected and standardized the average temperature (°C) and precipitation (mm) at a new location $\mathbf{s}^*$ for each spring. Then we estimated $\boldsymbol{\mu}(\mathbf{s}^*)$ with the posterior mean

$$\text{E}(\boldsymbol{\mu}(\mathbf{s}^*)|\mathbf{Y}) = \widehat{\boldsymbol{\mu}}(\mathbf{s}^*) \approx \frac{1}{K} \sum_{k=1}^{K} \text{softplus}(\mathbf{B}^{(k)}\mathbf{x}(\mathbf{s}^*) + \boldsymbol{\eta}^{(k)}(\mathbf{s}^*))$$

for each spring, where the (k) superscript denotes the kth MCMC iteration and $\boldsymbol{\eta}^{(k)}(\mathbf{s}^*)$ maximizes

$$[\boldsymbol{\eta}(\mathbf{s}^*)|\boldsymbol{\eta}^{(k)}] = \int \int [\boldsymbol{\eta}(\mathbf{s}^*), \boldsymbol{\phi}^{(k)}, \mathbf{C}^{(k)}|\boldsymbol{\eta}^{(k)}] d\boldsymbol{\phi}^{(k)} d\mathbf{C}^{(k)}.$$

Finally, we estimated the bioacoustic classification composition at $\mathbf{s}^*$ using (9) with $\widehat{\boldsymbol{\mu}}(\mathbf{s}^*)$ and $\widehat{\mathbf{V}}(\mathbf{s}^*) = f(\widehat{\boldsymbol{\mu}}(\mathbf{s}^*), \widehat{\boldsymbol{\gamma}})$, where $\widehat{\boldsymbol{\gamma}}$ is the posterior mean of $\boldsymbol{\gamma}$. We predicted acoustic compositions for each $\mathbf{s}^* \in \mathcal{S}$, where $\mathcal{S}$ is a fine grid spanning the study domain. The predicted acoustic compositions are plotted in Figure 5 for the springs of 2020–2023. Such predictions assume that the environmental effects on *D. pubescens* bioacoustic signal identification are time invariant. That is, the classification of *D. pubescens* acoustic behavior does not change in time and is only influenced by the local environment.
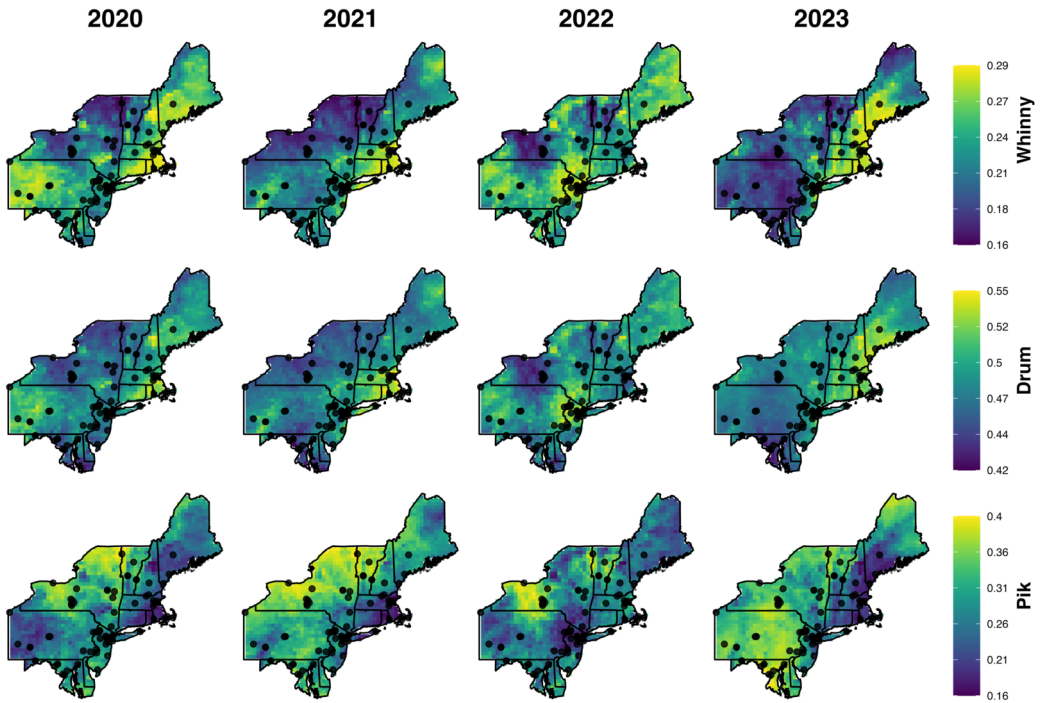
FIG. 5. *Predicted Dryobates pubescens bioacoustic signal compositions for the Northeastern United States during spring from* 2020–2023. *Dots denote observed locations.*

7.2. *Results.* The acoustic compositional predictions in Figure 5 are the predicted classifications of the bioacoustic signals—not the true signals themselves. Therefore, we make inference on bioacoustic classification, which can only serve as a proxy to the true acoustic behavior of *D. pubescens*. In general, drumming was the most commonly predicted bioacoustic behavior, whereas whinnying was the least. Whinnying was positively correlated with precipitation; see Figure 6 for maps of the average spring temperature (°C) and the average spring precipitation (mm). Northern New York experienced a relatively large amount of spring precipitation in 2022, resulting in higher predictive probabilities of whinnying. In 2023, northern Maine and Pennsylvania experienced less spring precipitation, resulting in lower predictive probabilities for whinnying. Additionally, the predictive probabilities of whinnying were affected by η, which captured environmental effects not captured by spring temperature and precipitation. In particular, the latent spatial effects for whinnying gave rise to slightly higher predictive probabilities in Maryland, New York, and Pennsylvania and slightly lower predictive probabilities in Massachusetts, New Hampshire, Rhode Island, and Vermont; see Appendix F for predictions of η across the study domain and a comparison between our proposed model and the same model without the spatial random effects η (Schwob et al. (2025)). The latter model performed nearly as well as our proposed model, indicating that most of the variation in the bioacoustic compositional predictions was captured by the environmental covariates.

Drumming classification was also positively correlated with precipitation. However, the predicted spatial patterns between whinnying and drumming were not identical, suggesting that drumming was also influenced by spring temperature and had a unique latent spatial field. Drumming was negatively correlated with spring temperature, which is most evident when comparing the northernmost and southernmost locations between whinnying and drumming classifications in Figure 5. Additionally, the latent spatial field for drumming resulted in slightly higher predictive probabilities in Maine, Maryland, Massachusetts, and Pennsylvania and slightly lower predictive probabilities in northern New York and Vermont. Finally, the
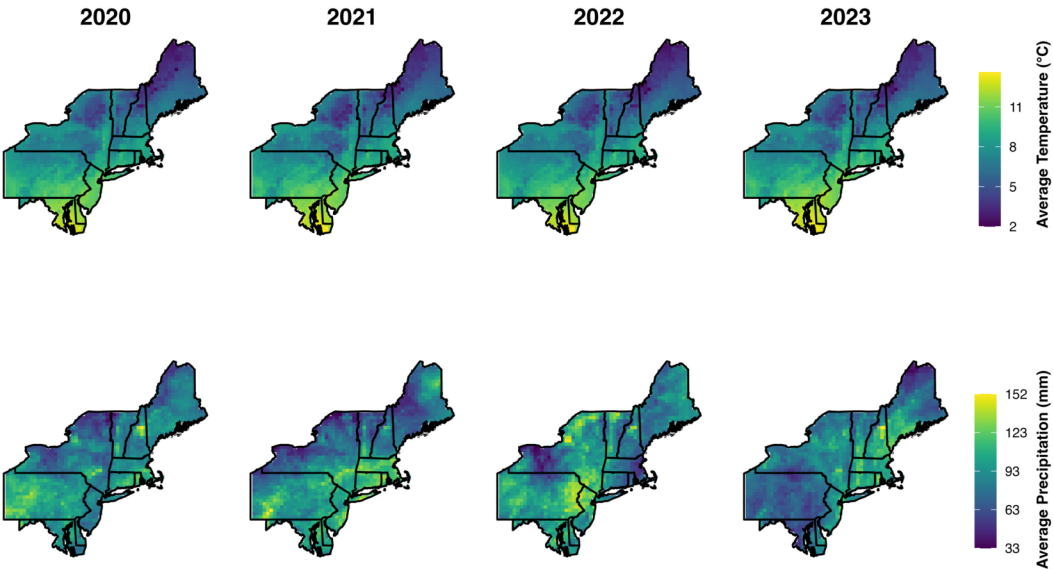
FIG. 6. *Average spring temperature (°C) and average spring precipitation (mm) in the Northeastern United States during spring from* 2020–2023.

spatial pattern of drumming classification was more consistent throughout the study period than the spatial pattern of whinnying classification, likely due to its dependence on spring temperature (which was more consistent than spring precipitation) and a smoother latent spatial field.

Piking prediction was the most variable each year, which we attribute to its negative correlation with spring precipitation. In particular, the compositional constraint required piking prediction to offset the positive correlation between whinnying/drumming and precipitation. The predictive probabilities of whinnying and drumming varied by roughly 13% each year and had similar spatial patterns. Due to the compositional constraint, the range of piking predictive probabilities varied up to 24% to offset the cumulative probability associated with whinnying and drumming classifications. To help satisfy the compositional constraint, the latent spatial field associated with piking classification was inversely related to the latent fields for whinnying and drumming classification.

In summary, drumming was the most commonly predicted bioacoustic signal during spring in the Northeastern United States. The spatial heterogeneity in predictive probabilities suggests that the bioacoustic behavior of *D. pubescens* was affected by the local environment. In warmer regions with high precipitation, we observed higher predictive probabilities for whinnying. In cool, wet regions, we observed an increase in predictive probabilities for drumming. In drier regions, we found higher predictive probabilities for piking.

Existing CoDA approaches are unable to obtain spatial compositional predictions, such as those depicted in Figure 5, for an arbitrary compositional data set. Our model allows for positive correlations between components and zero-valued components. Directional approaches for CoDA permit these features; however, existing directional methods may not render valid compositional predictions (i.e., predicting directional points outside the nonnegative orthant). Our model allows for zero-valued components, positive correlations, and is fully generative for compositional data, meaning that the predicted vector at every location in Figure 5 is a valid composition.

8. Conclusion. We developed a Bayesian hierarchical framework for spatial CoDA to analyze bioacoustic classifications. Compositional data are defined by two constraints: (i)

each component must be nonnegative, and (ii) the sum of the components must equal a constant (often taken to be 1). However, most existing methods for CoDA assume constraints that are stronger than necessary: (i) each component must be positive, and (ii) if the value of one component increases, the value of every other component must decrease (i.e., exclusively negative correlations). Our method naturally allows for zero-valued components and positive correlations between components. Therefore, our framework allows for the analysis of all compositional data without the need to manipulate or augment data (i.e., zero replacements, data imputation). Such features are appropriate when analyzing machine learning classifications, where positive correlations (between similar labels) and zero-valued components (due to confident predictions) may arise. In bioacoustic studies, passive acoustic monitoring (Keitt and Abelson (2021)) and deep supervised source separation (Sasek et al. (2024)) will increasingly render structural zeros.

Our framework features the first directional distribution truncated to the nonnegative orthant, which is exclusively where transformed compositional data reside. Standard directional distributions may include support outside the nonnegative orthant. In such cases, the distribution is not considered generative for compositional data because it may lead to directional observations that cannot be properly transformed back to the simplex. Existing directional approaches for compositional data either ignore this problem or circumvent it via folded distributions. Many directional distributions lack an analytically tractable normalizing constant, making folding a much more efficient approach than truncation. However, the recently proposed ESAG distribution has two key features: (i) it has an analytically tractable normalizing constant, and (ii) simulation is efficient due to its relationship with the multivariate Gaussian distribution. Thus, the truncation of the ESAG distribution is computationally feasible, and we find it more intuitive than the existing folding approaches.

Compositional data are an increasingly prevalent data source in spatial statistics (Feng, Zhu and Steen-Adams (2015)). However, previous directional approaches for compositional data have not accounted for spatial dependence. We proposed a spatial hyperspheric regression model that includes fixed and random spatial effects for the spatial analysis of bioacoustic signals. The latent spatial random effects were modeled as a multivariate spatial process because the relativity inherent in directional data suggests that the spatial effects for the components are interdependent. Although we used an LMC to model the multivariate spatial process, alternative approaches exist. Separable models are a special case of the LMC that offer computational advantages at the cost of specifying a single range parameter for all spatial fields (Schmidt and Gelfand (2003)). Graphical Gaussian processes construct cross-covariance functions from graphs to ensure process-level conditional independence between variables; in highly multivariate settings, these processes offer significant computational gains over the LMC (Dey, Datta and Banerjee (2022)). Additionally, functional graphical Gaussian processes have been developed for multivariate functional data (Dey, Datta and Banerjee (2022), Zapata, Oh and Petersen (2022)); Chandra et al. (2024) developed an autoregressive matrix-Gaussian copula graphical model to infer functional connectivity in auditory systems. Zhang and Banerjee (2022) proposed a matrix-normal approach for LMCs and spatial factor models in high-dimensional multivariate settings; Alie, Stephens and Schmidt (2024) proposed a matrix-normal LMC that induces sparsity through the introduction of structural zeros in the coregionalization matrix. Recently, Krock et al. (2023) wrote a multivariate Gaussian process as a linear combination of basis functions weighted with Gaussian graphical vectors, which scale well for highly multivariate nonstationary spatial data. Additionally, combining Hamiltonian Monte Carlo with a noncentral parameterization could improve the efficiency and mixing for high-dimensional multivariate spatial processes (Neal (2011)).

Machine learning classification provides an increasingly common source of compositional data. If the classification vectors are used for downstream analysis, then uncertainty from

the classification algorithm should be propagated to the subsequent model. We allowed for uncertainty propagation in the directional analysis of compositional data in the form of $\mathbf{1}_d \otimes \boldsymbol{\alpha}'$. The uncertainty propagation term in (8) did not affect the relative contribution of the components nor the orientation of dispersion; rather, it solely impacted the l_2-norm of $\boldsymbol{\mu}$, which governed the concentration of the distribution. Thus, the auxiliary information in $\mathbf{z}_i$ was used to propagate uncertainty in our downstream model. However, many compositional data sources are the result of observation rather than prediction and, consequently, do not necessitate uncertainty propagation. For example, community composition data measure the relative proportions of species in an ecosystem and are used to infer biodiversity, monitor environmental change, and inform conservation strategies (Jackson (1997)).

While the ESAG$^+$ density is feasible to evaluate, it is more computationally expensive than its nontruncated counterpart. Future work may include multistage Bayesian approaches, such as recursive Bayes, to decrease the computational costs of evaluating the ESAG$^+$ density (e.g., Hooten, Johnson and Brost (2021)). Other extensions to the present methodology might account for an autoregressive process on the regression coefficients or multivariate spatiotemporal dependencies in the data. Additionally, the concept of uncertainty propagation can be further explored in the context of compositional data, which is increasingly being generated by stochastic models.

We focused on the analysis of compositional data (i.e., bioacoustic classifications), although our methods can be adapted for unbounded directional data. For example, one only needs to set the standard ESAG distribution as the data model and use the identity link for $\boldsymbol{\mu}$. The rest of the framework remains the same to conduct a spatial directional data analysis featuring uncertainty propagation.

SUPPLEMENTARY MATERIAL

Web appendices (DOI: 10.1214/25-AOAS2082SUPPA; .pdf). Appendices containing (a) a list of the data used in the bioacoustic case study, (b) a description of our machine learning classifier, (c) the theoretical details of the alternative parameterization of the ESAG distribution (also see Yu and Huang (2024)), (d) a simulation study regarding the normalizing constant for the ESAG truncation, (e) our complete Bayesian hierarchical model, and (f) an extended discussion on the latent spatial random effects (Schwob et al. (2025)).

Code (DOI: 10.1214/25-AOAS2082SUPPB; .zip). The code used for analyses, which are also available at https://github.com/michaelschwob/Spatial-Hyperspheric-Models.

REFERENCES

AITCHISON, J. (1982). The statistical analysis of compositional data (with discussion). *J. Roy. Statist. Soc. Ser. B, Methodol.* **44** 139–177. MR0676206

ALIE, R., STEPHENS, D. A. and SCHMIDT, A. M. (2024). Computational considerations for the linear model of coregionalization. arXiv preprint. Available at arXiv:2402.08877.

BETANCOURT, I. and McLINN, C. M. (2012). Teaching with the Macaulay Library: An online archive of animal behavior recordings. *J. Microbiol. & Biol. Educ.* **13** 86–88.

BEZANSON, J., EDELMAN, A., KARPINSKI, S. and SHAH, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Rev.* **59** 65–98. MR3605826 https://doi.org/10.1137/141000671

BJERREGÅRD, M. B., MØLLER, J. K. and MADSEN, H. (2021). An introduction to multivariate probabilistic forecast evaluation. *Energy AI* **4** 100058.

CHANDRA, N. K., SITEK, K. R., CHANDRASEKARAN, B. and SARKAR, A. (2024). Functional connectivity across the human subcortical auditory system using an autoregressive matrix-Gaussian copula graphical model approach with partial correlations. *Imaging Neurosci.* **2** 1–23.

CRESSIE, N. and WIKLE, C. K. (2011). *Statistics for Spatio-Temporal Data. Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2848400

DENNIS, S. Y. III (1991). On the hyper-Dirichlet type 1 and hyper-Liouville distributions. *Comm. Statist. Theory Methods* **20** 4069–4081. MR1158563 https://doi.org/10.1080/03610929108830757

DEY, D., DATTA, A. and BANERJEE, S. (2022). Graphical Gaussian process models for highly multivariate spatial data. *Biometrika* **109** 993–1014. MR4519112 https://doi.org/10.1093/biomet/asab061

DODENHOFF, D. J. (2002). An analysis of acoustic communication within the social system of downy woodpeckers (*Picoides pubescens*) Ph.D. thesis The Ohio State Univ.

DOUMA, J. C. and WEEDON, J. T. (2019). Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression. *Methods Ecol. Evol.* **10** 1412–1430.

DRAPER, D. (1995). Assessment and propagation of model uncertainty (with discussion). *J. Roy. Statist. Soc. Ser. B, Methodol.* **57** 45–97. MR1325378

FARINA, A. and GAGE, S. H. (2017). *Ecoacoustics*: *The Ecological Role of Sounds*. Wiley, New York.

FENG, X., ZHU, J. and STEEN-ADAMS, M. M. (2015). On regression analysis of spatial proportional data with zero/one values. *Spat. Stat.* **14** 452–471. MR3431051 https://doi.org/10.1016/j.spasta.2015.07.007

GARCIA-FERNANDEZ, A., MASKELL, S., HORRIDGE, P. and RALPH, J. (2021). Gaussian tracking with Kent-distributed direction-of-arrival measurements. *IEEE Trans. Veh. Technol.* **70** 7249–7254.

GELFAND, A. E. and BANERJEE, S. (2010). Multivariate spatial process models. In *Handbook of Spatial Statistics*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 495–515. CRC Press, Boca Raton, FL. MR2730963 https://doi.org/10.1201/9781420072884-c28

GELFAND, A. E., SCHMIDT, A. M., BANERJEE, S. and SIRMANS, C. F. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization (with discussion). *TEST* **13** 263–312. MR2154003 https://doi.org/10.1007/BF02595775

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

GELFAND, A. E., SMITH, A. F. M. and LEE, T.-M. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *J. Amer. Statist. Assoc.* **87** 523–532. MR1173816

GLOOR, G. B. (2023). amIcompositional: Simple tests for compositional behaviour of high throughput data with common transformations. *Austrian J. Stat.* **52** 180–197.

GRANTHAM, N. S., GUAN, Y., REICH, B. J., BORER, E. T. and GROSS, K. (2020). MIMIX: A Bayesian mixed-effects model for microbiome data from designed experiments. *J. Amer. Statist. Assoc.* **115** 599–609. MR4107660 https://doi.org/10.1080/01621459.2019.1626242

GREENACRE, M. (2022). Compositional data analysis — linear algebra, visualization and interpretation. In *Innovations in Multivariate Statistical Modeling*: *Navigating Theoretical and Multidisciplinary Domains* 223–240. Springer, Berlin.

GREENACRE, M. (2024). The chiPower transformation: A valid alternative to logratio transformations in compositional data analysis. *Adv. Data Anal. Classif.* **18** 769–796. MR4798391 https://doi.org/10.1007/s11634-024-00600-x

GREENACRE, M., GRUNSKY, E., BACON-SHONE, J., ERB, I. and QUINN, T. (2023). Aitchison's compositional data analysis 40 years on: A reappraisal. *Statist. Sci.* **38** 386–410. MR4630375 https://doi.org/10.1214/22-sts880

HOETING, J. A., DAVIS, R. A., MERTON, A. A. and THOMPSON, S. E. (2006). Model selection for geostatistical models. *Ecol. Appl.* **16** 87–98.

HOOTEN, M. B., JOHNSON, D. S. and BROST, B. M. (2021). Making recursive Bayesian inference accessible. *Amer. Statist.* **75** 185–194. MR4256123 https://doi.org/10.1080/00031305.2019.1665584

JACKSON, D. A. (1997). Compositional data in community ecology: The paradigm or peril of proportions? *Ecology* **78** 929–940.

JOHNSON, S. G. (2013). QuadGK.jl: Gauss–Kronrod integration in Julia. Available at https://github.com/JuliaMath/QuadGK.jl.

KEITT, T. H. and ABELSON, E. S. (2021). Ecology in the age of automation. *Science* **373** 858–859.

KILHAM, L. (1962). Reproductive behavior of downy woodpeckers. *Condor* **64** 126–133.

KOSLOVSKY, M. D. (2023). A Bayesian zero-inflated Dirichlet-multinomial regression model for multivariate compositional count data. *Biometrics* **79** 3239–3251. MR4680718 https://doi.org/10.1111/biom.13853

KROCK, M. L., KLEIBER, W., HAMMERLING, D. and BECKER, S. (2023). Modeling massive highly multivariate nonstationary spatial data with the basis graphical lasso. *J. Comput. Graph. Statist.* **32** 1472–1487. MR4669262 https://doi.org/10.1080/10618600.2023.2174126

KYZYUROVA, K. N. (2017). On uncertainty quantification for systems of computer models. PhD thesis, Duke University.

LEININGER, T. J., GELFAND, A. E., ALLEN, J. M. and SILANDER, J. A. JR. (2013). Spatial regression modeling for compositional data with many zeros. *J. Agric. Biol. Environ. Stat.* **18** 314–334. MR3110896 https://doi.org/10.1007/s13253-013-0145-y

MAHAN, T. A. (1996). Analysis of the acoustic signals of adult male and female downy woodpeckers (Picoides pubescens) Ph.D. thesis Eastern Kentucky Univ.

MARTÍN-FERNÁNDEZ, J.-A., DAUNIS-I-ESTADELLA, J. and MATEU-FIGUERAS, G. (2015). On the interpretation of differences between groups for compositional data. *SORT* **39** 231–252. MR3467491

MCFEE, B., RAFFEL, C., LIANG, D., ELLIS, D. P., MCVICAR, M., BATTENBERG, E. and NIETO, O. (2015). librosa: Audio and music signal analysis in Python. *SciPy* 18–24.

MURPHY, K. P. (2022). *Probabilistic Machine Learning*: *An Introduction*. MIT Press, Cambridge.

MURRAY, I., ADAMS, R. and MACKAY, D. (2010). Elliptical slice sampling. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* 541–548.

NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 113–162. CRC Press, Boca Raton, FL. MR2858447

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* **29** 1350–1361.

PAINE, P. J., PRESTON, S. P., TSAGRIS, M. and WOOD, A. T. A. (2018). An elliptically symmetric angular Gaussian distribution. *Stat. Comput.* **28** 689–697. MR3761349 https://doi.org/10.1007/s11222-017-9756-4

PAINE, P. J., PRESTON, S. P., TSAGRIS, M. and WOOD, A. T. A. (2020). Spherical regression models with general covariates and anisotropic errors. *Stat. Comput.* **30** 153–165. MR4057477 https://doi.org/10.1007/s11222-019-09872-2

PAL, S. and HEUMANN, C. (2022). Clustering compositional data using Dirichlet mixture model. *PLoS ONE* **17** e0268438.

QUINN, T. P., ERB, I., RICHARDSON, M. F. and CROWLEY, T. M. (2018). Understanding sequencing data as compositions: An outlook and review. *Bioinformatics* **34** 2870–2878.

RODRÍGUEZ, A., DUNSON, D. B. and GELFAND, A. E. (2008). The nested Dirichlet process. *J. Amer. Statist. Assoc.* **103** 1131–1144. MR2528831 https://doi.org/10.1198/016214508000000553

SASEK, J., ALLISON, B., CONTINA, A., KNOBLES, D., WILSON, P. and KEITT, T. (2024). Semiautomated generation of species-specific training data from large, unlabeled acoustic datasets for deep supervised birdsong isolation. *PeerJ* **12** e17854.

SCEALY, J. L. and WELSH, A. H. (2011). Regression for compositional data by using distributions defined on the hypersphere. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 351–375. MR2815780 https://doi.org/10.1111/j.1467-9868.2010.00766.x

SCEALY, J. L. and WELSH, A. H. (2014). Fitting Kent models to compositional data with small concentration. *Stat. Comput.* **24** 165–179. MR3165546 https://doi.org/10.1007/s11222-012-9361-5

SCEALY, J. L. and WOOD, A. T. A. (2019). Scaled von Mises–Fisher distributions and regression models for paleomagnetic directional data. *J. Amer. Statist. Assoc.* **114** 1547–1560. MR4047280 https://doi.org/10.1080/01621459.2019.1585249

SCEALY, J. L. and WOOD, A. T. A. (2023). Score matching for compositional distributions. *J. Amer. Statist. Assoc.* **118** 1811–1823. MR4646608 https://doi.org/10.1080/01621459.2021.2016422

SCHMIDT, A. M. and GELFAND, A. E. (2003). A Bayesian coregionalization approach for multivariate pollutant data. *J. Geophys. Res., Atmos.* **108**.

SCHWOB, M. R., HOOTEN, M. B., CALZADA, N. M. and KEITT, T. H. (2025). Supplement to "Spatial Hyperspheric Models for Compositional Data." https://doi.org/10.1214/25-AOAS2082SUPPA, https://doi.org/10.1214/25-AOAS2082SUPPB

SHORT, L. L. and SANDSTRÖM, G. F. (1982). Woodpeckers of the world. Monograph Series of the Delaware Museum of Natural History.

STEVENS, S. S., VOLKMANN, J. and NEWMAN, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *J. Acoust. Soc. Amer.* **8** 185–190.

STEWART, C. (2017). An approach to measure distance between compositional diet estimates containing essential zeros. *J. Appl. Stat.* **44** 1137–1152. MR3638436 https://doi.org/10.1080/02664763.2016.1193846

STEWART, C., LANG, S. L. C., IVERSON, S. and BOWEN, W. D. (2022). Measuring repeatability of compositional diet estimates: An example using quantitative fatty acid signature analysis. *Ecol. Evol.* **12** e9428.

SUEUR, J. and FARINA, A. (2015). Ecoacoustics: The ecological investigation and interpretation of environmental sound. *Biosemiotics* **8** 493–502.

TANG, Y., MA, L. and NICOLAE, D. L. (2018). A phylogenetic scan test on a Dirichlet-tree multinomial model for microbiome data. *Ann. Appl. Stat.* **12** 1–26. MR3773384 https://doi.org/10.1214/17-AOAS1086

TIPTON, J. R., HOOTEN, M. B., NOLAN, C., BOOTH, R. K. and MCLACHLAN, J. (2019). Predicting paleoclimate from compositional data using multivariate Gaussian process inverse prediction. *Ann. Appl. Stat.* **13** 2363–2388. MR4037434 https://doi.org/10.1214/19-aoas1281

TSAGRIS, M. and STEWART, C. (2018). A Dirichlet regression model for compositional data with zeros. *Lobachevskii J. Math.* **39** 398–412. MR3789428 https://doi.org/10.1134/s1995080218030198

WANG, F. and GELFAND, A. E. (2013). Directional data analysis under the general projected normal distribution. *Stat. Methodol.* **10** 113–127. MR2974815 https://doi.org/10.1016/j.stamet.2012.07.005

WIEMANN, P. F. V., KNEIB, T. and HAMBUCKERS, J. (2024). Using the softplus function to construct alternative link functions in generalized linear models and beyond. *Statist. Papers* **65** 3155–3180. MR4767659 https://doi.org/10.1007/s00362-023-01509-x

XIE, J. and ZHU, M. (2022). Sliding-window based scale-frequency map for bird sound classification using 2d-and 3d-CNN. *Expert Syst. Appl.* **207** 118054.

XU, Y., KONG, Q., WANG, W. and PLUMBLEY, M. D. (2018). Large-scale weakly supervised audio classification using gated convolutional neural network. In 2018 *IEEE International Conference on Acoustics*, *Speech and Signal Processing* 121–125.

YU, S., DRTON, M. and SHOJAIE, A. (2024). Interaction models and generalized score matching for compositional data. In *Learning on Graphs Conference* 20–1. PMLR.

YU, Z. and HUANG, X. (2024). A new parameterization for elliptically symmetric angular Gaussian distributions of arbitrary dimension. *Electron. J. Stat.* **18** 301–334. MR4699571 https://doi.org/10.1214/23-ejs2210

YU, Z. and HUANG, X. (2025). Regression analysis of elliptically symmetric directional data. *Comput. Statist. Data Anal.* **208** Paper No. 108167. MR4874241 https://doi.org/10.1016/j.csda.2025.108167

ZAPATA, J., OH, S. Y. and PETERSEN, A. (2022). Partial separability and functional graphical models for multivariate Gaussian processes. *Biometrika* **109** 665–681. MR4472841 https://doi.org/10.1093/biomet/asab046

ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99** 250–261. MR2054303 https://doi.org/10.1198/016214504000000241

ZHANG, L. and BANERJEE, S. (2022). Spatial factor modeling: A Bayesian matrix-normal approach for misaligned data. *Biometrics* **78** 560–573. MR4450576 https://doi.org/10.1111/biom.13452