Title
Bayesian Regression for Dependent Tensor-Valued Data from Exponential Families

Journal
Journal of Agricultural, Biological and Environmental Statistics

ISSN
1085-7117

Authors
North, Joshua S
Risser, Mark D
Rhoades, Alan M

Publication Date
2026-07-28

DOI
10.1007/s13253-026-00748-3

Supplementary materials for this article are available at https://doi.org/10.1007/s13253-026-00748-3.

Bayesian Regression for Dependent Tensor-Valued Data from Exponential Families

Joshua S. NORTH, Mark D. RISSER, and Alan M. RHOADES

In the Earth system sciences, identifying physically meaningful patterns from dependent, high-dimensional, tensor-valued data is important for understanding variability in the Earth system. Orthonormal matrices are fundamental for statistical modeling of dependent data due to their role in matrix and tensor decomposition methods. While significant effort has been devoted to statistical inference for dependent matrix-valued data, these developments have not yet been extended to tensor-valued data. Here, we propose a general framework for dependent tensor-valued generalized linear mixed models (GLMMs) with a non-separable and non-stationary structured random effect. The random effect is parameterized using a tensor decomposition consisting of structured random orthonormal matrices, whose structures and associated coefficients are inferred during model estimation. Fully Bayesian inference enables joint estimation of all components of the dependent random effect and fixed effect within a unified statistical model. We validate our method with two synthetic examples and apply it to two real-world applications within atmospheric science, highlighting its capacity to uncover structured modes of variability across space, time, and vertical level. More broadly, our method provides a path forward for identifying structured patterns and their relative importance with Bayesian uncertainty quantification for arbitrary dependent, exponential family-distributed tensor-valued data.

Key Words: Basis Functions; Bayesian Hierarchical Model; Bayesian Tensor Decomposition; Low-Rank Representation; Orthonormal Matrix; Tensor-Valued Generalized Linear Mixed Model.

1. INTRODUCTION

In the Earth sciences, weather and climate variables are often indexed over multiple dimensions, including space, time, vertical level, and realization (or ensemble member). Such tensor-valued data are therefore stored and organized in tensors within network Common Data Form (netCDF) files, which are standard across widely-used reanalysis (e.g., Hersbach et al., 2020) and climate model (e.g., Eyring et al., 2016) databases. There are

---

Joshua S. North (✉), Mark D. Risser, Alan M. Rhoades, Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA (E-mail: *jsnorth@lbl.gov*).

Journal of Agricultural, Biological, and Environmental Statistics
https://doi.org/10.1007/s13253-026-00748-3

Published online: 28 July 2026

two features that make statistical modeling of tensor-valued climate data challenging. First, the physical properties of the Earth system induce dependence within and across different dimensions of the tensor, and it is critical to account for these structures in a statistical analysis. Second, while some variables are approximately Gaussian-distributed, other variables, such as precipitation, relative humidity, and wind speed, are non-Gaussian. Earth scientists' ability to extract physically interpretable patterns from tensor-valued climate data relies on flexible tensor decompositions that explicitly model dependence in high-dimensional, non-Gaussian measurements.

Tensor decompositions can be thought of as the higher-order, or N-dimensional, extension of singular value decomposition (SVD) or principal component analysis (PCA). These decompositions factor tensor-valued data into weights and orthonormal matrices (Kolda 2006; Kolda and Bader 2009), which simultaneously summarize key characteristics of the data and enable dimension reduction for data compression and storage. Two common tensor decompositions are the Tucker (Tucker 1966) and CANDECOMP/PARAFAC (CP; Carroll and Chang 1970; Harman and Harman 1976; Kiers 2000), which are frequently used in medical imaging, signal processing, ecology, climate sciences, and many other scientific fields (Sidiropoulos et al. 2017; Chatzichristos et al. 2019; Korevaar et al. 2020; Shivadekar et al. 2023). Despite their broad importance in the climate sciences and beyond, little attention has been devoted to statistical estimation and uncertainty quantification of the individual components in the tensor decomposition.

For matrix-valued space-time data, basis functions from *matrix* decompositions are commonly used to summarize spatial and temporal features of the data. For example, empirical orthogonal functions (EOFs; see Lorenz 1956; North et al. 1982; Hannachi et al. 2007) are widely used to summarize modes of sea surface temperature variability (Thompson and Wallace 2000; Mantua and Hare 2002) and quantify human-induced changes to surface air temperature (O'Brien and Deser 2023). Statistical modeling of climate data often incorporates spatial and temporal information via spatially-indexed basis functions with temporal coefficients or via space-time random effects within a generalized linear mixed-model (GLMM Banerjee et al. 2015; Hughes and Haran 2013). The GLMM is generally more suitable than the generalized linear model because it accounts for structured variation not accounted for by the fixed effects. When considering tensor data with more than two dimensions, the nature of the random effect in a GLMM depends on what the higher dimensions represent. Multivariate space-time methods for three-dimensional tensors over *space* $\times$ *time* $\times$ *process/variable* typically consider a very small number of slices in the process dimension (e.g., five or fewer variables) and focus on explicit modeling of cross-covariances via co-regionalization (Gelfand et al. 2004; Finley et al. 2008) or multivariate covariance functions (Genton and Kleiber 2015). Very little work has been done for random effects modeling of three (or more)-dimensional tensors from a single variable, e.g., geopotential height defined over *space* $\times$ *time* $\times$ *pressure level*, and none have focused on explicitly inferring the basis functions defined over the tensor's dimensions.

Traditionally, the basis functions employed in the space, time, or space-time random effect are pre-specified and the associated coefficients are modeled as random variables. This is generally done for two reasons. First, in order to ensure that the basis functions describe independent patterns, they must be orthonormal and therefore collectively form an orthonormal matrix. Orthonormal matrices are contained in the Stiefel manifold $\mathcal{V}_{k,n} = \{\mathbf{X} \in \mathbb{R}^{n\times k} : \mathbf{X}'\mathbf{X} = \mathbf{I}_k\}$ where $\mathbf{I}_k$ is the identity matrix, and specifying a distribution on the Stiefel manifold can be challenging (Mardia and Jupp 1999; Chikuse 2003). Second, being able to parameterize and sample from a distribution on the Stiefel manifold can be difficult. Recently, multiple approaches for specifying and sampling from distributions for orthonormal matrices in a Bayesian framework have been proposed, along with various techniques to incorporate the distributions into matrix factorization models such as for SVD, PCA, and factor analysis (Hoff 2007, 2009; Byrne and Girolami 2013; Pourzanjani et al. 2021; Jauch et al. 2021, 2025; North et al. 2024). However, little work has focused on incorporating these advancements to tensor-valued data.

Motivated by the use of GLMMs for spatio-temporal data, we propose a general framework for tensor-valued GLMMs. The fixed effect is motivated from prior work on generalized linear models for tensor-valued data (Zhou et al. 2013; Guhaniyogi et al. 2015; Lock 2018; Guhaniyogi and Spencer 2021; Billio et al. 2023; Wang and Xu 2024; Lei et al. 2024). The random effect borrows ideas from the spatio-temporal literature and is parameterized by a low-rank process (Cressie and Wikle 2011; Kang and Cressie 2011; Wikle and Zammit-Mangion 2019). In contrast to traditional spatio-temporal models, the basis functions comprising the low-rank random effect are estimated and can be modeled dependently. Our proposed model is a tensor extension of the statistical models for matrix decompositions and enables a probabilistic representation of the CP-decomposition, whereby all basis functions and weights in the decomposition can be modeled in a probabilistic framework. Importantly, we show the resulting covariance of the random effect process is non-separable and non-stationary, making it well-suited for complex environmental data.

The remainder of the manuscript is organized as follows. Section 2 presents the parameterization of the tensor-valued random effect along with tensor notation that will underlie the tensor-valued GLMM. Section 3 presents the tensor-valued GLMM, discusses various choices for modeling the fixed effect, dependent random effect, and how to perform model estimation. The general model is applied to two synthetic examples in Section 4, one assuming Gaussian distributed data and one assuming Gamma distributed data, illustrating the model's flexibility and ability to recover important features of the data. In Section 5, we demonstrate the method's ability to infer modes of variability across multiple dimensions through two real-world applications: (1) atmospheric pressure data surrounding a major weather blocking event over the United States, and (2) specific humidity data over the Southwest United States to identify spatial patterns of environmental drivers. Section 6 concludes the paper, discussing limitations and possible future directions.

2. DEPENDENT TENSOR-VALUED RANDOM EFFECT

In spatio-temporal modeling, a random effect is used to characterize any residual spatial or temporal structure not accounted for by the fixed effects. These dependence structures are commonly modeled via a reduced rank matrix decomposition, resulting in a low-rank random effect that summarizes the spatial and temporal dependence in the data. When data are instead indexed over more than two dimensions (e.g., space, time, vertical level, ensemble member, ...), mathematical notation benefits from the use of tensors, the higher-order analog of matrices. Motivated by the low-rank random effect models in the spatial statistics literature, we propose a dependent random effect for tensor-valued random variables using the CANDECOMP/PARAFAC decomposition (CP; Carroll and Chang 1970; Harman and Harman 1976; Kiers 2000). Importantly, our framework involves inference for both the basis functions and core tensor within the CP decomposition.

This section provides a road map of the building blocks for our model for a structured, tensor-valued random variable. We begin with a brief overview of tensor notation and decompositions in Section 2.1 (see Supplement S.2 for additional details). Then, we describe a prior distribution for the components of the tensor decomposition that allows us to encode structure or dependence within the basis functions (Section 2.2). Last, we discuss properties of our proposed tensor-valued random effect (Section 2.3).

2.1. TENSOR NOTATION AND DECOMPOSITION

Define $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \dots \times I_N}$ to be a N-dimensional tensor process that is indexed over $\mathbf{s}^{(k)} = \{s_1^{(k)}, \dots, s_{I_k}^{(k)}\} \in \mathbb{D}^{(k)}$ where $\mathbb{D}^{(k)}$ is some domain for $k = 1, \dots, N$. For example, below we consider two applications using atmospheric data indexed by space, time, and vertical pressure level. In both of these settings, $\mathcal{Y}$ is a third-order tensor with $N = 3$ where $\mathbf{s}^{(1)}$ are spatial locations in some spatial domain $\mathbb{D}^{(1)} \subset \mathbb{R}^2$, $\mathbf{s}^{(2)}$ are discrete times recorded over some continuous time window $\mathbb{D}^{(2)} = [T_{start}, T_{end}] \subset \mathbb{R}^+$, and $\mathbf{s}^{(3)}$ are discrete vertical levels $\mathbb{D}^{(3)} \subset \mathbb{R}^+$.

A tensor can be converted to a matrix by *matricization* (also known as unfolding or flattening), where the elements of the tensor are reordered into a matrix. The n-mode matricization of the tensor $\mathcal{Y}$, denoted by $\mathbf{Y}_{(n)}$, arranges the mode$-n$ fibers (the higher-order equivalent of matrix rows and columns) to be columns in the resulting matrix. For example, some possible modes of $\mathcal{Y}$ are $\mathbf{Y}_{(1)} \in \mathbb{R}^{I_1 \times (I_2 \times \dots \times I_N)}$, $\mathbf{Y}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \times \dots \times I_{n-1} \times I_{n+1} \times \dots \times I_N)}$, and so on. To multiply a tensor by a matrix $\mathbf{B} \in \mathbb{R}^{J \times I_n}$, we use the n-mode product (i.e., multiply a tensor by a matrix or vector in mode n). The n-mode product of the tensor $\mathcal{Y}$ and matrix $\mathbf{B}$ is denoted as $\mathcal{Y} \times_n \mathbf{B}$ and is of size $I_1 \times I_2 \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N$. Equivalently, in terms of unfolded (matricized) tensors, $\mathcal{Z} = \mathcal{Y} \times_n \mathbf{B} \Leftrightarrow \mathbf{Z}_{(n)} = \mathbf{B}\mathbf{Y}_{(n)}$. Working with tensors requires the notion of outer products, denoted by $\circ$, the Kronecker product, denoted $\otimes$, and the Khatri-Rao product, denoted by $\odot$. The Khatri-Rao product of $\mathbf{B} \in \mathbb{R}^{I \times J}$ and $\mathbf{C} \in \mathbb{R}^{M \times J}$ is defined as

$$\mathbf{D} = \mathbf{B} \odot \mathbf{C} = \left[\mathbf{b}_1 \otimes \mathbf{c}_1 \ \mathbf{b}_2 \otimes \mathbf{c}_2 \ \cdots \ \mathbf{b}_J \otimes \mathbf{c}_J\right], \qquad \mathbf{D} \in \mathbb{R}^{IM \times J}.$$

The concept of matrix decompositions can be extended to tensors, termed tensor decomposition, wherein a tensor is factored into the product of tensors and matrices. Similar to matrix decomposition methods, there are many different types of tensor decomposition methods, e.g., De Lathauwer et al. (2000); Vasilescu and Terzopoulos (2002); Kapteyn et al. (1986); Tucker (1966); Carroll and Chang (1970); Harman and Harman (1976); Kiers (2000). Our proposed model employs the CP decomposition which results in a unique solution for the underlying basis functions and core tensor (unlike, e.g., the Tucker decomposition; Tucker

1966; Kolda and Bader 2009). Denote the $(i_1, i_2, ..., i_N)$ element of $\mathcal{Y}$ as $y_{i_1,i_2,...,i_N}$. The reduced rank CP decomposition of $\mathcal{Y}$ is defined as $y_{i_1,i_2,...,i_N} = \sum_{r=1}^{R} g_r a_{i_1 r}^{(1)} a_{i_2 r}^{(2)} \cdots a_{i_N r}^{(N)}$, or equivalently

$$\mathcal{Y} = \sum_{r=1}^{R} g_r \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \cdots \circ \mathbf{a}_r^{(N)}, \tag{2.1}$$

where $R \le \min(I_1, \dots, I_N)$ is the rank, $\mathbf{G} = \mathrm{diag}(g_1, ..., g_R)$ is a diagonal matrix of basis coefficients with $g_1 > g_2 > ... > g_R$, and $\mathbf{A}^{(k)} \in \mathbb{R}^{I_k \times R}$ is an orthonormal matrix where each column $r = 1, ..., R$ is given by $\mathbf{a}_r^{(k)} \equiv (a_{1r}, ..., a_{I_k r})$. The mode-$k$ matricized version of the reduced rank CP decomposition is

$$\mathbf{Y}_{(k)} = \mathbf{A}^{(k)} \mathbf{G} (\mathbf{A}^{(N)} \odot \cdots \odot \mathbf{A}^{(k+1)} \odot \mathbf{A}^{(k-1)} \odot \cdots \odot \mathbf{A}^{(1)})'.$$

2.2. PRIORS FOR THE STRUCTURED ORTHONORMAL BASIS FUNCTIONS AND WEIGHTS

As is common within spatio-temporal statistical modeling, we assume the random effect $\mathcal{Y}$ is parameterized using a basis expansion, which we define using the CP decomposition (2.1). Importantly, if $\mathcal{Y}$ is assumed to have structure, or dependence, the individual bases $\mathbf{a}_r^{(k)}$ should also be dependently structured, and we refer to $\mathbf{A}^{(k)} = [\mathbf{a}_1^{(k)}, \dots, \mathbf{a}_R^{(k)}]$ as a structured orthonormal matrix. Generally speaking, we treat the collection of orthonormal matrices $\{\mathbf{A}^{(k)}\}_{k=1,...,N}$ and weights in the expansion as random variables to be estimated. Additionally, each basis function (column) in the matrix will have a unique dependence structure, which we parameterize using kernel functions with estimable hyperparameters. Due to this construction, the random effect $\mathcal{Y}$ is a non-stationary and non-separable process, which we discuss in detail below.

To estimate structured orthonormal matrices in a Bayesian fashion, North et al. (2024) proposed a method based on the projected normal (PN) distribution, which we briefly summarize here. Let $\mathbf{A}_{-r}^{(k)} = [\mathbf{a}_1^{(k)}, \dots, \mathbf{a}_{r-1}^{(k)}, \mathbf{a}_{r+1}^{(k)}, \dots, \mathbf{a}_R^{(k)}]$ and $\mathbf{G}_{-r} = \mathrm{diag}(g_1, ..., g_{r-1}, g_{r+1}, ..., g_R)$. Then $\mathbf{a}_r^{(k)} | \mathbf{A}_{-r}^{(k)} \overset{d}{=} \mathbf{N}_{-r}^{(k)} \widetilde{\mathbf{a}}_r^{(k)} | \mathbf{A}_{-r}^{(k)}$, where the columns of the $I_k \times (I_k - R + 1)$ matrix $\mathbf{N}_{-r}^{(k)}$ form an orthonormal basis for the null space of $\mathbf{A}_{-r}^{(k)}$. The projected weight function $\widetilde{\mathbf{a}}_r^{(k)}$ satisfies $\widetilde{\mathbf{a}}_r^{(k)} | \mathbf{A}_{-r}^{(k)} \sim \mathrm{PN}_{I_k - R+1}(\mathbf{0}, \mathbf{N}_{-r}^{(k)\prime} \boldsymbol{\Omega}_r^{(k)} \mathbf{N}_{-r}^{(k)})$, where PN denotes the projected normal distribution and $\boldsymbol{\Omega}_r^{(k)} \in \mathbb{R}^{I_k \times I_k}$ is the associated covariance matrix of $\widetilde{\mathbf{a}}_r^{(k)}$, where $\boldsymbol{\Omega}_r^{(k)}$ controls the dependence structure of $\mathbf{a}_r^{(k)}$ (e.g., $\boldsymbol{\Omega}_r^{(k)} = \mathbf{I}_{I_k}$ results in no structure and independent elements). To make the PN distribution tractable for sampling, the distribution is augmented with a latent length variable, which in this context is g_r. The resulting joint distribution $(g_r, \widetilde{\mathbf{a}}_r^{(k)}) | \mathbf{A}_{-r}^{(k)}$ is

$$p(g_r, \widetilde{\mathbf{a}}_r^{(k)} | \mathbf{A}_{-r}^{(k)}) = \frac{g_r^{I_k - R} \exp\left\{-\frac{1}{2}(g_r \widetilde{\mathbf{a}}_r^{(k)})' (\mathbf{N}_r^{(k)\prime} \boldsymbol{\Omega}_r^{(k)} \mathbf{N}_r^{(k)})^{-1} (g_r \widetilde{\mathbf{a}}_r^{(k)})\right\}}{(2\pi)^{-(I_k - R+1)/2} |\mathbf{N}_r^{(k)\prime} \boldsymbol{\Omega}_i \mathbf{N}_i|^{-1/2}} \mathbb{I}(\widetilde{\mathbf{a}}_r^{(k)} \in \mathcal{V}_{1, I_k - R+1}),$$

where $\mathcal{V}_{1,I_k-R+1}$ denotes the Stiefel manifold (in this context this is equivalent to the $I_k - R$ dimensional unit sphere $\mathbb{S}^{I_k-R}$). We denote this as $g_r\widetilde{\mathbf{a}}_r^{(k)}|\mathbf{A}_{-r}^{(k)} \sim \text{SPN}_{I_k-R+1}(\boldsymbol{\Omega}_r^{(k)})$, the *structured projected normal* (SPN) distribution, which has one parameter $\boldsymbol{\Omega}_r^{(k)}$ that controls the structure of the basis function.

To model dependence within the bases $\mathbf{a}_r^{(k)}$, $\boldsymbol{\Omega}_r^{(k)}$ can be parameterized as $\boldsymbol{\Omega}_r^{(k)} = \gamma_r^{2(k)}\mathbf{C}_r^{(k)}$, where $\mathbf{C}_r^{(k)}$ is a positive-definite correlation matrix that is modeled a using kernel function and $\gamma_r^{2(k)}$ is a scaling parameter related to g_r. We use this parameterization and model the elements of $\mathbf{C}_r^{(k)}$ using kernel functions $C^{(k)}(\mathbf{s}^{(k)}; \boldsymbol{\theta}_r^{(k)})$ where $\boldsymbol{\theta}_r^{(k)}$ are estimable hyperparameters and $\mathbf{s}^{(k)}$ are as defined above. For all examples presented here, we use either the Matérn, Gaussian, or identity kernel. For the Matérn kernel, $\boldsymbol{\theta}_r^{(k)} = \{\nu_r^{(k)}, \rho_r^{(k)}\}$ represent the differentiability and length-scale of the implied stochastic process, and for the Gaussian kernel, $\boldsymbol{\theta}_r^{(k)} = \{\rho_r^{(k)}\}$ is the length-scale of the implied stochastic process. For the Matérn kernel ν is assumed fixed, and for both the Matérn and Gaussian kernels we estimate the length-scale parameter ρ, thereby enabling each basis function to have a different covariance structure.

2.3. Properties of the Random Effect

To determine key properties of $\mathcal{Y}$, we begin by providing a statement of its covariance matrix (proof deferred to the Supplement S.1).

Proposition 1. *When $g_r\widetilde{\mathbf{a}}_r^{(k)}|\mathbf{A}_{-r}^{(k)} \sim SPN_{I_k-R+1}(\boldsymbol{\Omega}_r^{(k)})$, the covariance of $\mathcal{Y} = \sum_{r=1}^R g_r\mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \cdots \circ \mathbf{a}_r^{(N)}$ can be written as the product-sum of mode-separable stationary covariance matrices that are a function of $\boldsymbol{\Omega}_r^{(k)} = \gamma_r^{2(k)}\mathbf{C}_r^{(k)}$, where $\mathbf{C}_r^{(k)}$ the correlation matrix for the rth basis function of mode k and is parameterized using a kernel function $C^{(k)}(\mathbf{s}^{(k)}; \boldsymbol{\theta}_r^{(k)})$ with estimable hyperparameters $\boldsymbol{\theta}_r^{(k)}$.*

The covariance of $\mathcal{Y}$ is non-stationary and non-separable because it is the product-sum of R distinct stationary covariance (correlation) matrices operating on different scales (Fuentes 2001; Fuentes et al. 2008; Banerjee et al. 2015; Porcu et al. 2021). Because we allow each covariance matrix to have dimension specific and basis-function specific hyperparameters, we cannot factor the covariance matrix of $\mathcal{Y}$ into the product of mode-specific covariance matrices, hence it is non-separable. Similarly, we are summing multiple stationary processes operating on different scales, hence it is non-stationary. Alternatively, if all the basis-function specific hyperparameters were the same, $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}_r^{(k)}$ for all $r = 1, \ldots, R$, then every basis function within mode k would have the same covariance matrix $C^{(k)}(\mathbf{s}^{(k)}; \boldsymbol{\theta}^{(k)}) \equiv C^{(k)}(\mathbf{s}^{(k)}; \boldsymbol{\theta}_r^{(k)})$ and the covariance of $\mathcal{Y}$ would be separable and stationary. This would result in a covariance formulation similar to separable models where $\text{cov}(\mathbf{Y}_{(N)}) = \boldsymbol{\Sigma}_N \otimes \cdots \otimes \boldsymbol{\Sigma}_1$ with $\boldsymbol{\Sigma}_k$ a covariance for the kth dimension. Additionally, because we are able to estimate the basis functions $\mathbf{a}_r^{(k)}$ comprising $\mathcal{Y}$, we are able to make statements about the different components (e.g., the r basis function of mode k loads positively/negatively on region x and negatively/positively on region y).

We summarize the three important properties of $\mathcal{Y}$ as follows:

Property 1. *The* $\text{cov}(\mathcal{Y})$ *is non-separable*: Determined using proposition 1;

Property 2. *The* $\mathrm{cov}(\mathcal{Y})$ *is non-stationary*: Determined using proposition 1;

Property 3. $\mathcal{Y}$ *contains interpretable components*: By construction we estimate the individual basis functions $\mathbf{a}_r^{(k)}$.

These three properties yield a very flexible and interpretable random effect for dependent high-dimensional data.

3. GENERAL PROBABILISTIC MODEL

As discussed in the introduction, environmental data often have multiple dimensions, such as space, time, vertical/pressure level, and ensemble member. The data may have measurable covariates that are indexed over the same dimensions or a subset of the dimensions (e.g., indexed only over space or time but constant for other dimensions). This necessitates the ability to relate covariates with the same or fewer dimensional properties to the variable of interest through a fixed effect. However, since the fixed effect will never explain all of the variability in the data we use the dependently structured tensor-valued random effect described in Section 2 to account for the remaining structured variation. Here, we present our tensor-valued generalized linear mixed model (GLMM). Notably, the GLMM utilizes the dependently structured tensor-valued random effect defined in Section 2 to estimate dominant patterns of variability from the process of interest.

Define $\mathcal{Z} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ to be a tensor of observed data with individual element $z_i \equiv z_{i_1,i_2,\ldots,i_N}$ that is also indexed over $\mathbf{s}^{(k)} = \{s_1^{(k)}, \ldots s_{I_k}^{(k)}\} \in \mathbb{D}^{(k)}$ where $\mathbb{D}^{(k)}$ is some domain for $k = 1, \ldots, N$. We assume z_i is a random variable whose probability density function is a member of the exponential family,

$$f(z_i|\phi_i, \tau_i) = \exp\left\{\frac{z_i\phi_i - b(\phi_i)}{a(\tau_i)} + c(z_i, \tau_i)\right\},$$

where ϕ_i and $\tau_i > 0$ denote the natural and dispersion parameters, respectively, and $a(\cdot), b(\cdot)$, and $c(\cdot)$ are known functions. The GLMM relates a fixed effect m_i, correlated random effect y_i, and uncorrelated random effect ϵ_i to the conditional mean $\mu_i = E[z_i|m_i, y_i, \epsilon_i]$ through the monotonic link function $h_i = g(\mu_i) = m_i + y_i + \epsilon_i$.

We define our latent process as

$$\mathcal{H} = \mathcal{M} + \mathcal{Y} + \mathcal{E}, \tag{3.1}$$

where $\mathcal{M}$ denotes the fixed effect, $\mathcal{Y} = \sum_{r=1}^{R} g_r \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \cdots \circ \mathbf{a}_r^{(N)}$ the structured random effect, and $\mathcal{E}$ the unstructured random effect. We assume each element $\epsilon_{i_1,i_2,\cdots,i_N} \overset{\text{i.i.d}}{\sim} \mathrm{Gau}(0, \sigma^2)$, resulting in $\mathcal{E} \sim \mathrm{GTD}_{I_1 \times I_2 \ldots \times I_N}(0, \sigma^2)$, where GTD denotes the Gaussian tensor distribution (see Supplement S.3 for details). Note, care must be taken to ensure the model is identifiable. For example, if z_i are assumed to be Bernoulli distributed and the probit link is used, σ^2 is not be identifiable and customarily set equal to one (Albert and Chib 1993). In Sections 4 and 5, we explore the scenarios where z_i is Gaussian and Gamma distributed, and discuss the steps to verify identifiability for both settings.

3.1. MODEL FOR THE FIXED EFFECT

Tensor regression for fixed effects has been well studied (Zhou et al. 2013; Guhaniyogi et al. 2015; Lock 2018; Guhaniyogi and Spencer 2021; Billio et al. 2023; Wang and Xu 2024; Lei et al. 2024), and can refer to predicting a scalar output from a tensor, matrix output from a tensor, or tensor output from a either a scalar, matrix, or tensor. Here, we focus on estimating tensor-valued output using a combination of a scalar, matrix, or tensor input using two approaches, either with standard matrix methods and matching dimensions or using tensor methods. We discuss the implementation of both.

3.1.1. Fixed Effect: Matrix Covariates

Define the vectorized version of (3.1) as

$$\mathbf{h} = \mathbf{m} + \mathbf{y} + \boldsymbol{\epsilon}$$

where each element is vectorized by stacking the fibers of the tensor along one dimension where now $\mathbf{h}, \mathbf{m}, \mathbf{y}, \boldsymbol{\epsilon} \in \mathbb{R}^L$ with $L = \prod_{k=1}^N I_k$. For simplicity, assume the data are vectorized along the first dimension, resulting in e.g.,

$$\mathbf{h} = [h_{(1,1,\ldots,1)}, h_{(2,1,\ldots,1)}, \ldots, h_{(I_1,1,\ldots,1)}, h_{(1,2,\ldots,1)}, \ldots, h_{(I_1,I_2,\ldots,I_N)}]'.$$

This formulation allows for standard regression techniques, such as $\mathbf{m} = \mathbf{X}\boldsymbol{\beta}$ where $\mathbf{X} \in \mathbb{R}^{L \times P_m}$ is a matrix of measurable covariates and $\boldsymbol{\beta}$ is a length$-P_m$ vector of parameters. Note, care needs to be taken to ensure the rows of $\mathbf{X}$ are properly aligned with their corresponding rows in the vectorized latent process $\mathbf{h}$, and subsequently the data $\mathcal{Z}$.

3.1.2. Fixed Effect: Tensor Covariates

Let $\mathcal{V}^{(p)} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, $p = 1, \ldots, P_t$ be a tensor of measurable covariates and $\mathcal{B}^{(p)} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ the associated parameter tensor. Define the linear model for $\mathcal{M}$ in (3.1) as

$$\mathcal{M} = \sum_{p=1}^{P_t} \mathcal{V}^{(p)} * \mathcal{B}^{(p)}. \tag{3.2}$$

In practice, $\mathcal{B}^{(p)}$ is over-parameterized and instead of estimating it directly, it is modeled using a Tucker decomposition (see Supplement S.2 for definition).

Lemma 1. *Let $\mathcal{Y} = \mathcal{X} * \mathcal{B}$ with $\mathcal{Y}, \mathcal{X}, \mathcal{B} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ and $*$ the element-wise multiplication of two equal dimension tensors (also known as the Hadamard product). Representing $\mathcal{B}$ using the Tucker decomposition, $\mathcal{B} \approx \mathcal{D} \times_1 \boldsymbol{\Psi}^{(1)} \times_2 \boldsymbol{\Psi}^{(2)} \cdots \times_N \boldsymbol{\Psi}^{(N)}$ where $\mathcal{D} \in \mathbb{R}^{J_1 \times \ldots \times J_N}$ and $\boldsymbol{\Psi}^{(k)} \in \mathbb{R}^{I_k \times J_k}$. Then*

$$vec(\mathbf{Y}_{(k)}) = \boldsymbol{\Gamma}\boldsymbol{\delta},$$

where $\mathbf{\Gamma} = (vec(\mathbf{X}_{(k)}) \odot \mathbf{1}') * \widetilde{\mathbf{\Gamma}}$ *with* $\mathbf{1}$ *a length-*$\prod_{k=1}^{N} J_k$ *vector of ones,* $\widetilde{\mathbf{\Gamma}} = (\mathbf{\Psi}^{(N)} \otimes \ldots \otimes \mathbf{\Psi}^{(k+1)} \otimes \mathbf{\Psi}^{(k-1)} \otimes \ldots \otimes \mathbf{\Psi}^{(1)}) \otimes \mathbf{\Psi}^{(k)}$*, and* $\boldsymbol{\delta} = vec(\mathbf{D}_{(k)})$*. Note, the choice of matricizing over the* k^{th} *dimension is arbitrary. Proof deferred to Supplement S.1.*

From Lemma 1, the vectorized equivalent of (3.2) is

$$\mathbf{m}_{(k)} \equiv \text{vec}(\mathbf{M}_{(k)}) = \sum_{p=1}^{P_t} \mathbf{\Gamma}^p \boldsymbol{\delta}^{(p)}, \tag{3.3}$$

where $\mathbf{\Gamma}^{(p)} = (\text{vec}(\mathbf{V}_{(k)}^{(p)}) \odot \mathbf{1}')\widetilde{\mathbf{\Gamma}}^{(p)}$ and $\widetilde{\mathbf{\Gamma}}^{(p)} = (\mathbf{\Psi}^{(N)(p)} \otimes \ldots \otimes \mathbf{\Psi}^{(1)(p)})$. The matrices $\mathbf{\Psi}^{(k)(p)} \in \mathbb{R}^{I_k \times J_k^{(p)}}$ for $k = 1, \ldots, N$ and $p = 1, \ldots, P_t$ with $J_k^{(p)} \ll I_k$ are pre-specified fixed basis functions specific to each dimension k and do not need to be the same for every p. Parameterizing $\mathcal{M}$ in this manner results in estimating $\boldsymbol{\delta}^{(p)}$ instead of $\mathcal{B}^{(p)}$, reducing the number of parameters from $\sum_{p=1}^{P} \prod_{k=1}^{N} I_k$ to $\sum_{p=1}^{P} \prod_{k=1}^{N} J_k^{(p)}$, and providing the ability to incorporate dependence into $\mathcal{B}^{(p)}$. The resulting vectorized form of (3.1) is

$$\text{vec}(\mathbf{H}_{(k)}) \equiv \mathbf{h}_{(k)} \sim \text{N}\left(\sum_{p=1}^{P_t} \mathbf{\Gamma}^p \boldsymbol{\delta}^{(p)} + \text{vec}(\mathbf{Y}_{(k)}), \mathbf{\Phi}_{(k)} \otimes \mathbf{\Sigma}^{(k)}\right).$$

3.1.3. Fixed Effect: Matrix and Tensor Covariates

Combining the fixed effect based on matrix (3.1.1) and tensor (3.1.2) covariates,

$$\text{vec}(\mathbf{H}_{(k)}) \equiv \mathbf{h}_{(k)} \sim \text{N}\left(\text{vec}((\mathbf{X}\boldsymbol{\beta})_{(k)}) + \sum_{p=1}^{P} \mathbf{\Gamma}^p \boldsymbol{\delta}^{(p)} + \text{vec}(\mathbf{Y}_{(k)}), \mathbf{\Phi}_{(k)} \otimes \mathbf{\Sigma}^{(k)}\right), \tag{3.4}$$

where $\mathbf{X}$ are the matrix covariates with corresponding length-p_M vector of parameters $\boldsymbol{\beta}$, $(\mathbf{X}\boldsymbol{\beta})_{(k)}$ denotes the orientation of the k-fold matricization of the matrix-based fixed effect, and $\sum_{p=1}^{P} \mathbf{\Gamma}^p \boldsymbol{\delta}^{(p)}$ comprise of the tensor-based fixed effect. In practice, we use the mode-1 matricization (e.g., $k = 1$) and then vectorize over that transformation, but the method is agnostic to the matricization mode.

3.2. MODEL ESTIMATION

Inference is conducted using a Bayesian framework by employing Markov chain Monte Carlo (MCMC). To complete our model specification, we assign priors to the model parameters: the fixed effect matrix and tensor parameters $\boldsymbol{\beta}$ and $\boldsymbol{\delta}^{(p)}$, respectively; the structured random effect parameters g_r and $\mathbf{a}_r^{(k)}$; the unstructured variance σ; and any parameters from the data distribution.

For both $\boldsymbol{\beta}$ and $\boldsymbol{\delta}^{(p)}$, we assign diffuse normal priors $\boldsymbol{\beta} \sim \text{N}_{p_M}(\mathbf{0}, \sigma_\beta^2 \mathbf{I}_{p_M})$ and $\boldsymbol{\delta}^{(p)} \sim \text{N}_{p_T}(\mathbf{0}, \sigma_\delta^2 \mathbf{I}_{p_T})$, where σ_β^2 and σ_δ^2 are set to 100 by default. For σ^2, we assign the non-informative half-t prior on the standard deviation $\sigma \sim Half - t(1, A)$, with A large (e.g., $A \sim 1 \times 10^5$; see Huang and Wand 2013, fordetailsanddiscussion). The priors for $\boldsymbol{\beta}$, $\boldsymbol{\delta}^{(p)}$, and σ are conjugate and allow for Gibbs steps of these parameters (see Supplement S.6).

For the random effect parameters, we use the structured projected normal (SPN) distribution discussed in Section 2.2, which requires (3.1) to have the conditional form $\mathbf{a}_r^{(k)}|\mathbf{A}_{-r}^{(k)}$. Taking the k-fold matricization of (3.1),

$$\begin{aligned}\mathbf{H}_{(k)} &= \mathbf{M}_{(k)} + \mathbf{A}^{(k)}\mathbf{G}\Big(\mathbf{A}^{(N)} \odot \mathbf{A}^{(N-1)} \cdots \odot \mathbf{A}^{(k+1)} \odot \mathbf{A}^{(k-1)} \cdots \odot \mathbf{A}^{(1)}\Big)' + \mathbf{E}_{(k)} \\ &= \mathbf{M}_{(k)} + \mathbf{A}^{(k)}\mathbf{G}\mathbf{B}^{(k)} + \mathbf{E}_{(k)},\end{aligned} \tag{3.5}$$

where $\mathbf{G} = \text{diag}(g_1, ..., g_R)$, the Khatri-Rao product of the structured orthonormal matrices without the k matrix as $\mathbf{B}^{(k)} \equiv \Big(\mathbf{A}^{(N)} \odot \mathbf{A}^{(N-1)} \cdots \odot \mathbf{A}^{(k+1)} \odot \mathbf{A}^{(k-1)} \cdots \odot \mathbf{A}^{(1)}\Big)'$, and the k-fold orientated unstructured random effect $\mathbf{E}_{(k)} \sim \text{MN}_{I_k \times L_k}(\mathbf{0}, \mathbf{\Sigma}^{(k)}, \mathbf{\Phi}^{(k)})$ where $\mathbf{\Phi}^{(k)} = \mathbf{\Sigma}^{(N)} \otimes \cdots \otimes \mathbf{\Sigma}^{(k+1)} \otimes \mathbf{\Sigma}^{(k-1)} \cdots \otimes \mathbf{\Sigma}^{(1)}$ with $\mathbf{\Sigma}^{(k)} = \mathbf{I}_{I_k}$ the identity matrix. The k-folded distribution (3.5) can be rearranged (see Supplement S.4 for details) to yield

$$\begin{aligned}&p(\mathbf{H}_{(k)}|\mathbf{M}_{(k)}, \mathbf{a}_r^{(k)}, \mathbf{b}_r^{(k)}, g_r, \mathbf{A}_{-r}^{(k)}, \mathbf{B}_{-r}^{(k)}, \mathbf{G}_{-r}, \mathbf{\Sigma}^{(k)}, \mathbf{\Phi}^{(k)}) = \\ &\frac{1}{(2\pi)^{I_k L_k/2}|\mathbf{\Sigma}^{(k)}|^{L_k/2}|\mathbf{\Phi}^{(k)}|^{I_k/2}} \times \exp\Big\{-\frac{1}{2}\text{tr}\Big[\mathbf{\Phi}^{(k)^{-1}}\mathbf{D}_{-r}^{(k)\prime}\mathbf{\Sigma}^{(k)^{-1}}\mathbf{D}_{-r}^{(k)} - \\ &2g_r\mathbf{a}_r^{(k)\prime}\mathbf{\Sigma}^{(k)^{-1}}\mathbf{D}_{-r}^{(k)}\mathbf{\Phi}^{(k)^{-1}}\mathbf{b}_r^{(k)\prime} + g_r^2\mathbf{a}_r^{(k)\prime}\mathbf{\Sigma}^{(k)^{-1}}\mathbf{a}_r^{(k)}\mathbf{b}_r^{(k)}\mathbf{\Phi}^{(k)^{-1}}\mathbf{b}_r^{(k)\prime}\Big]\Big\},\end{aligned} \tag{3.6}$$

where $\mathbf{b}_r^{(k)} = \mathbf{a}_r^{(N)} \odot \cdots \odot \mathbf{a}_r^{(k+1)} \odot \mathbf{a}_r^{(k-1)} \cdots \odot \mathbf{a}_r^{(1)}$, $\mathbf{B}_{-r}^{(k)} \equiv \Big(\mathbf{A}_{-r}^{(N)} \odot \mathbf{A}_{-r}^{(N-1)} \cdots \odot \mathbf{A}_{-r}^{(k+1)} \odot \mathbf{A}_{-r}^{(k-1)} \cdots \odot \mathbf{A}_{-r}^{(1)}\Big)'$, and $\mathbf{D}_{-r}^{(k)} = \mathbf{H}_{(k)} - \mathbf{M}_{(k)} - \mathbf{A}_{-r}^{(k)}\mathbf{G}_{-r}\mathbf{B}_{-r}^{(k)}$. Importantly, we now have the conditional form $\mathbf{a}_r^{(k)}|\mathbf{A}_{-r}^{(k)}$. Using the substitution $\mathbf{a}_r^{(k)}|\mathbf{A}_{-r}^{(k)} \overset{d}{=} \mathbf{N}_{-r}^{(k)}\widetilde{\mathbf{a}}_r^{(k)}|\mathbf{A}_{-r}^{(k)}$, the resulting full conditional distribution for $\widetilde{\mathbf{a}}_r^{(k)}$ is

$$\begin{aligned}&[\widetilde{\mathbf{a}}_r^{(k)}|\cdot] \sim N(S^{-1}m, S^{-1})\mathbb{I}\Big(\widetilde{\mathbf{a}}_r^{(k)} \in \mathcal{V}_{1,I_k-r+1}\Big), \\ &\quad m = g_r\mathbf{N}_r^{(k)\prime}\mathbf{\Sigma}^{(k)^{-1}}\mathbf{D}_{-r}^{(k)}\mathbf{\Phi}^{(k)^{-1}}\mathbf{b}_r^{(k)\prime}, \\ &\quad S^{-1} = g_r^2\Big[\Big(\mathbf{b}_r^{(k)}\mathbf{\Phi}^{(k)^{-1}}\mathbf{b}_r^{(k)\prime}\Big)\mathbf{N}_r^{(k)\prime}\mathbf{\Sigma}^{(k)^{-1}}\mathbf{N}_r^{(k)} + \Big(\mathbf{N}_r^{(k)\prime}\mathbf{\Omega}_r^{(k)}\mathbf{N}_r^{(k)}\Big)^{-1}\Big].\end{aligned} \tag{3.7}$$

To produce a sample of $\mathbf{a}_r^{(k)}$, first sample $\widetilde{\mathbf{a}}_r^{(k)}$ from (3.7) then set $\mathbf{a}_r^{(k)} = \mathbf{N}_{-r}^{(k)}\widetilde{\mathbf{a}}_r^{(k)}$. This prior is convenient because $\mathbf{a}_r^{(k)}$ can be sampled using a Gibbs update step.

The core tensor $\mathbf{G} = \text{diag}(g_1, \ldots, g_R)$ from the CP-decomposition does not enable the use of a conjugate prior. Instead, we estimate $\mathbf{G}$ using a Metropolis-Hastings (MH) step. Other MCMC sampling methods could be used, but we found MH to be adequate. Using the reformulated distribution (3.6), the posterior distribution of $g_r|\mathbf{G}_{-r}$ is

$$L(g_r|\mathbf{H}_{(k)}, \boldsymbol{\theta}) = \prod_{k=1}^{N} p(\mathbf{H}_{(k)}|g_r, \boldsymbol{\theta})p(g_r\widetilde{\mathbf{a}}_r^{(k)}|\mathbf{A}_{-r}^{(k)})$$

where $\boldsymbol{\theta} = \{\mathbf{M}_{(k)}, \mathbf{a}_r^{(k)}, \mathbf{b}_r^{(k)}, \mathbf{A}_{-r}^{(k)}, \mathbf{B}_{-r}^{(k)}, \mathbf{G}_{-r}, \boldsymbol{\Sigma}^{(k)}, \boldsymbol{\Phi}^{(k)}\}$. We use the log of the posterior for the MH updates, given as

$$\begin{aligned}\mathcal{L}(g_r|\mathbf{H}_{(k)}, \boldsymbol{\theta}) \equiv \log L(g_r|\mathbf{H}_{(k)}, \boldsymbol{\theta}) &= \sum_{k=1}^{N} \log p(\mathbf{H}_{(k)}|g_r, \boldsymbol{\theta}) + \log p(g_r \widetilde{\mathbf{a}}_r^{(k)}|\mathbf{A}_{-r}^{(k)}) \qquad (3.8)\\ &\propto \sum_{k=1}^{N} g_r \mathbf{a}_r^{(k)\prime} \boldsymbol{\Sigma}^{(k)^{-1}} \mathbf{D}_{-r}^{(k)} \boldsymbol{\Phi}^{(k)^{-1}} \mathbf{b}_r^{(k)\prime} - \frac{1}{2} g_r^2 \mathbf{a}_r^{(k)\prime} \boldsymbol{\Sigma}^{(k)^{-1}} \mathbf{a}_r^{(k)} \mathbf{b}_r^{(k)} \boldsymbol{\Phi}^{(k)^{-1}} \mathbf{b}_r^{(k)\prime}\\ &\quad + (I_k - R)\log(g_r) - \frac{1}{2} g_r^2 \widetilde{\mathbf{a}}_r^{(k)\prime} \left(\mathbf{N}_r^{(k)\prime} \boldsymbol{\Omega}_r^{(k)} \mathbf{N}_r^{(k)}\right)^{-1} \widetilde{\mathbf{a}}_r^{(k)}.\end{aligned}$$

Our full MCMC sampling scheme utilizes a combination of Gibbs, MH, and Hamiltonian Monte Carlo (HMC; Neal 2011) update steps, and is included in the Supplement S.6 along with additional implementation details and full conditional distributions.

4. SYNTHETIC EXAMPLES

We conduct two simulation studies to illustrate general cases where the data are assumed to be (1) Gaussian and (2) gamma distributed. For both simulations, we utilize the full latent model with both matrix and tensor fixed effects and a structured random effect. Using the Gaussian simulation, we make a brief comparison to a currently accepted methods for modeling dependent high-dimensional data.

The latent process for both examples is simulated as follows. We simulate $I_1 = 50$, $I_2 = 50$, and $I_3 = 100$ points evenly spaced over the domain $(0, 10) \times (-10, 10) \times (0, 30)$, resulting in $\mathcal{H}$ containing 250000 elements. For the matrix fixed effects, we let $P_m = 4$ with $\boldsymbol{\beta} = (-1, -0.6, 0.5, 1)$ and the elements of $\mathbf{X} \overset{\text{i.i.d}}{\sim} \text{Gau}(0, \text{var}(\mathcal{Y}))$. For the tensor fixed effects, we take $P_t = 2$ where the elements of the tensor covariates $\mathcal{V}^{(p)} \overset{\text{i.i.d}}{\sim} \text{Gau}(0, \text{var}(\mathcal{Y}))$. For the basis functions in the tensor fixed effect $\{\boldsymbol{\Psi}^{(k)(p)}\}_{k=1,\ldots,N}$ we use Wendland basis functions with three knot points spaced equally across each respective domain, and $\boldsymbol{\delta}^{(p)} \overset{\text{i.i.d}}{\sim} \text{Gau}(0, 1)$, resulting in the tensor fixed effect having structured coefficients $\mathcal{B}^{(p)}$. The random effect $\mathcal{Y}$ is specified to be of rank $R = 6$ with $\mathbf{g} = (90, 80, 70, 60, 50, 40)$. For the orthonormal matrices $\mathbf{A}^{(k)}$, we specify a Matérn correlation with smoothness $\nu = 3.5$ and length-scale $\rho_1 = \{4, 4, 3, 3, 2, 2\}$, $\rho_2 = \{3, 4, 5, 6, 7, 8, 4\}$, and $\rho_3 = \{10, 9, 8, 6, 7, 5\}$ for each dimension, respectively. The generating mechanism for the orthonormal matrices is described in North et al. (2024). This ensures each basis function *and* each dimension have a different spatial ranges. Last, σ^2, the process variance, is drawn from a half-Cauchy with location parameter set to 1. For all simulations, we obtain 10000 posterior draws treating the first 9000 as burnin.

4.1. GAUSSIAN RESPONSE

When the data are Gaussian distributed, the data model simplifies to

$$\mathcal{Z} = \mathcal{M} + \mathcal{Y} + \mathcal{E} \qquad (4.1)$$

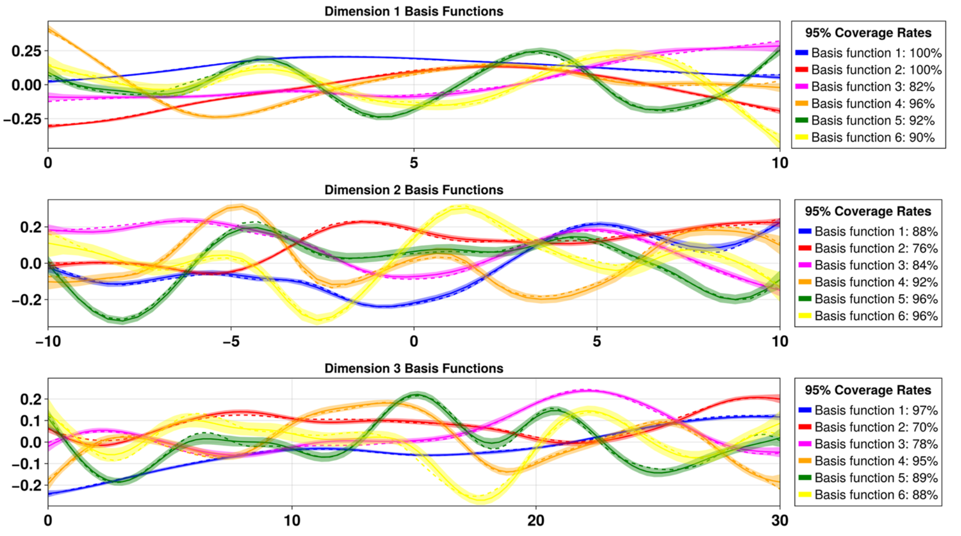

Figure 1. Basis function results for the gaussian simulation. Posterior mean (solid line), 95% credible interval (shaded region), and true (dashed line) basis functions for each simulated dimension from 1 (top) to 3 (bottom). The legend caption on the right indicates the 95% coverage rate .

Here, we restrict each covariance matrix to be diagonally structured; $\mathbf{\Sigma}^{(k)} = \text{diag}(\sigma_1^{(k)}, \ldots, \sigma_{I_k}^{(k)})$, although this can be relaxed if other model assumptions are made (see e.g., Hoff 2011). Even with the diagonal restriction, the covariance matrices are not identifiable (see e.g. the discussion of non-identifiability for the covariance matrices from the matrix normal distribution by Glanz and Carvalho 2018). Define $\mathbf{\Phi} = \mathbf{\Sigma}^{(N)} \otimes \mathbf{\Sigma}^{(N-1)} \otimes \cdots \otimes \mathbf{\Sigma}^{(1)}$. For any constants c_1, ..., $c_N \neq 0$ such that $c_1 c_2 \cdots c_N = 1$, $c_N \mathbf{\Sigma}^{(N)} \otimes c_{N-1} \mathbf{\Sigma}^{(N-1)} \otimes \cdots \otimes c_1 \mathbf{\Sigma}^{(1)} = \mathbf{\Phi}$. To ensure identifiability for the matrix normal distribution, Glanz and Carvalho (2018) fix the scale of the covariance matrices and separately estimate the scale of the overall variance. Extending this to an N-order system, we propose the following amendment to the covariance specification, $\mathbf{\Phi} = c \times \mathbf{\Sigma}^{(N)} \otimes \mathbf{\Sigma}^{(N-1)} \otimes \cdots \otimes \mathbf{\Sigma}^{(1)}$ where c is the scale of the overall variance and require $\sigma_1^{2(k)} = 1$ for $k = 1, ..., N$. Note, the choice of the top-left entry is arbitrary. An alternate parameterization is discussed in Hoff (2011), but we do not explore this choice. To simulate data from this model, the only adjustment is now $\sigma_{i_k}^{2(k)}$ for $k = 1, \ldots, N$ and $i_k = 2, \ldots, I_k$ are drawn from a truncated half-Cauchy that is bounded between 0.8 and 1.2 and with location parameter set to 1. We choose to bound the variance parameters so the resulting $\mathbf{\Phi}$ is on approximately the same scale as the fixed and random effects and no term dominates the signal. Additional details on the GTD and how to simulate from a GTD are included in the supplement.

The basis function true (dashed line), posterior mean (solid line), and 95% credible intervals (CI) are shown in Figure 1 along with their point-wise 95% coverage rate. We see the structure of each basis function is recovered and the first basis function for each dimension has less uncertainty than the preceding basis function, which corresponds to the strength of the signal (e.g., value of g). Additionally, Figure S.1 shows three slices chosen at random, one for each dimension (top to bottom), of the posterior mean (left column), true

Table 1. Basis-function averaged point-wise RMSE for each matrix $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ and the random effect surface $\mathcal{Y}$ computed from our approach and both local, Vecchia-approximated GP (lvGP) scenarios

Dimension	Our approach	lvGP: Scenario (1)	lvGP: Scenario (2)
$\mathbf{A}_1$	0.0105	0.0573	0.0591
$\mathbf{A}_2$	0.0112	0.0469	0.0480
$\mathbf{A}_3$	0.0089	0.0378	0.0490
$\mathcal{Y}$	0.04260	0.1675	0.1676

(middle column), and posterior difference (right column) of the random effect. In addition to capturing the individual basis functions for each dimension, we are able to recover the full latent process. For the structured component of the fixed effect, Figure S.2 show six slices chosen at random, one for each dimension (left to right) and for each covariate (top/bottom) of the true (black line), posterior mean (blue line), and 95% CI (shaded region). Again, we are able to capture the parameters and visually have reasonable coverage. Due to the number of parameters in the model, we are unable to visualize all combinations. Instead, Table S.1 summarizes the 95% coverage rates for all parameters, and we see reasonable coverage rates.

For comparison, we apply a local, Vecchia-approximated Gaussian process (lvGP) to estimate the structured random effect, wherein we separately fit Vecchia-approximated stationary GPs (Vecchia 1988; Katzfuss and Guinness 2021) to subsets of the data. Local GPs are used to both reduce the computational burden of fitting GPs and to capture nonstationarity (stationary within each local region but non-stationary over the whole domain); we furthermore use Vecchia approximations since even subsets of our data are too large for training exact GPs. Full details of our analysis are deferred to the Supplement S.8. The lvGP is used to recover the basis functions and random effect under two different scenarios: (1) assuming a random effect only model (e.g., no $\mathcal{M}$) and (2) a two stage approach where the fixed effect is first estimated and then the local GP is fit on the residuals after removing the estimated fixed effect. We obtain lvGP basis functions by computing the CP-decomposition of the associated lvGP predictions. The estimated basis functions from scenario (1) and (2) and their uncertainty are shown in Figures S.3 and S.4. In both cases we see that while the GP approach captures the general trends, it fails to recover the mean (95% prediction intervals do not overlap the truth) and the estimates are much noisier than the truth. We also computed the basis-function averaged point-wise root mean square error (RMSE) for each matrix and the random effect surface in Table 1. Our approach results in a 4-5 fold decrease in the RMSE relative to lvGP, and is able to recover the basis function estimates (overlapping 95% CIs in Figure 1).

4.2. GAMMA RESPONSE

The Fisher Information matrix for the shape-scale Gamma distribution has non-zero off-diagonal elements, implying the shape θ and scale k are correlated. Reparameterizing the

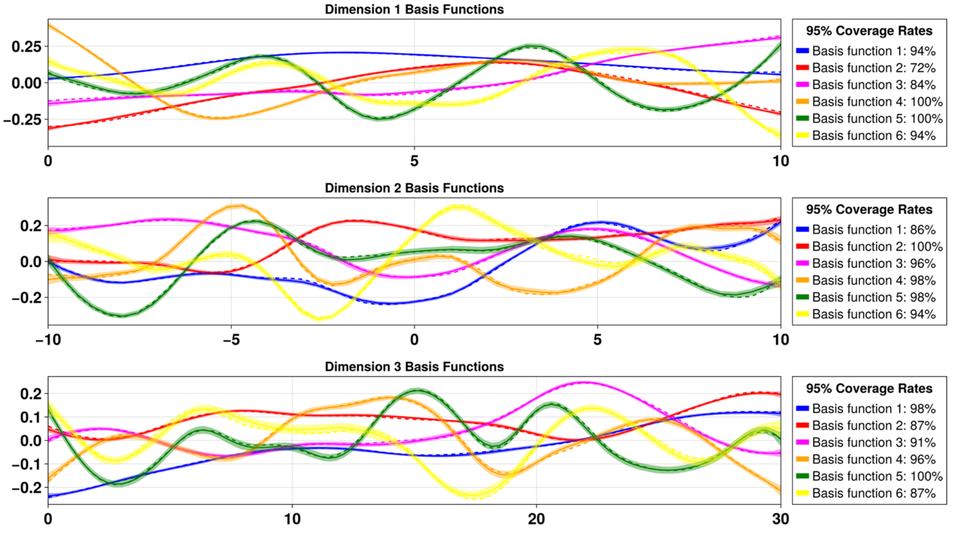

Figure 2. Basis function results for the gamma simulation. Posterior mean (solid line), 95% credible interval (shaded region), and true (dashed line) basis functions for each simulated dimension from 1 (top) to 3 (bottom). The legend caption on the right indicates the 95% coverage rate .

distribution where $\eta = k\theta$ and $\xi = k$, the resulting density is

$$f(x|\eta, \xi) = \frac{1}{\Gamma(\xi)\left(\frac{\eta}{\xi}\right)^{\xi}} x^{\xi-1} e^{-x\xi/\eta}, \tag{4.2}$$

where $E(x) = \eta$ and $\text{var}(x) = \eta^2/\xi$ and the off-diagonal elements of the Fisher Information matrix are zero. We denote the reparameterized Gamma as $X \sim \text{R-Gamma}(\eta, \xi)$ and a shape-scale Gamma random variable as $X \sim \text{Gamma}(k, \theta)$. For all examples presented here, we parameterize Gamma distributed random variables using (4.2).

To simulate data arising from the R-Gamma distribution, we simulate the latent process $h_{i_1,i_2,\ldots,i_N}$ as described at the start of Section 4. We set $\xi = 9$, define $\theta_{i_1,i_2,\ldots,i_N} = \exp\{h_{i_1,i_2,\ldots,i_N}\}/\xi$, and simulate $z_{i_1,i_2,\ldots,i_N} \sim \text{Gamma}(\theta_{i_1,i_2,\ldots,i_N}, \xi)$. For parameter estimation using the R-Gamma distribution, we employ the log-link function, e.g. $E(z_{i_1,i_2,\ldots,i_N}) = \log(\eta_{i_1,i_2,\ldots,i_N}) = h_{i_1,i_2,\ldots,i_N}$. For the scale ξ, we assign the Half-Cauchy$(0, \infty)$ prior with scale 1 and use the HMC updating scheme.

While we used local GPs for comparison in Section 4.1, it is less straightforward to do so for non-Gaussian data (requiring either a data transformation or sampling of a latent GP). Given the relatively poor performance of GPs for modeling tensor-valued random effects for Gaussian-distributed data and the dearth of off-the-shelf methods for non-Gaussian tensor-valued data, we do not attempt a comparison for this example.

Results from applying our proposed method to the gamma simulation are shown in Figure 2: basis function ground truth (dashed line), posterior mean (solid line), and 95% credible intervals (CI) along with their point-wise 95% coverage rate. As with the Gaussian simulation, we see the structure of each basis function is recovered and the first basis function for each dimension has less uncertainty than the last basis function. Additionally, Figure S.5 shows the same three random slices as the Gaussian simulation of the random effect, where again we are able to recover the latent process in addition to the individual basis functions. For the structured component of the fixed effect, Figure S.6 show the same six random slices as for the Gaussian simulation, where we are able to capture the parameters and have reasonable coverage. Table S.2 summarizes the 95% coverage rates for all parameters, where we see reasonable coverage rates.

5. APPLICATIONS

We now apply the proposed method to two different tensor-valued atmospheric processes: (1) atmospheric pressure and (2) specific humidity. In each case, the tensor is defined over space $\times$ time $\times$ vertical level. The first application considers a random effect only model and is the model-based analogue for CP decomposition. The second application is more involved and exemplifies our proposed method's ability to identify major modes of variability for non-Gaussian data with a more complex set of fixed effects. For both applications, we obtain 10000 posterior draws and treat the first 9000 as burnin. All computation was performed in Julia (Bezanson et al. 2017) using a 2021 Apple MacBook Pro with an Apple M1 Pro chip (8-core CPU), 16 GB RAM, running macOS Sonoma 14.5. Total compute time was just under 17 and 11 hours for the atmospheric pressure and specific humidity examples, respectively.

5.1. ATMOSPHERIC PRESSURE: GAUSSIAN DISTRIBUTED DATA

We analyze daily average atmospheric pressure taken from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis fifth generation (ERA5) dataset (Hersbach et al. 2020) at 50 millibars (mb) intervals from 950mb to 50mb between February 1^{st} to 28^{th} 2021. This time period corresponds to an atmospheric blocking event, which formed around February 6th and dissipated on the 21st. The blocking event resulted in a major cold air outbreak across the United States, leaving many people without power, resulting in over 100 fatalities, and breaking multiple cold temperature records (Davis et al. 2022). We consider only locations over land contained within the geographic region spanning 125°W to 65°W and 25°N to 50°N. Data are re-gridded to 1° and centered and scaled by pressure level (Figure S.7), resulting in 577,752 data points and the data being approximately Gaussian. We model the data using (4.1), but do not include any covariates (e.g., $\mathbf{X}$ or $\boldsymbol{\Gamma}$), resulting in a random effect only model with rank $R = 10$. The rank was chosen based on the first 10 basis functions from the CP-decomposition explaining approximately 80% of the variation in the data. For the basis function correlation structures, we assume a Matérn correlation structure over space with a smoothness parameter of 3.5 and a Gaussian correlation structure for both time and pressure level.

The first six estimated basis functions for space, time, and pressure-level are shown in Figure 3 A, B, and C, respectively. Here, we elect to only show the first six basis functions because the remaining basis functions show little to no spatial pattern, and their corre-

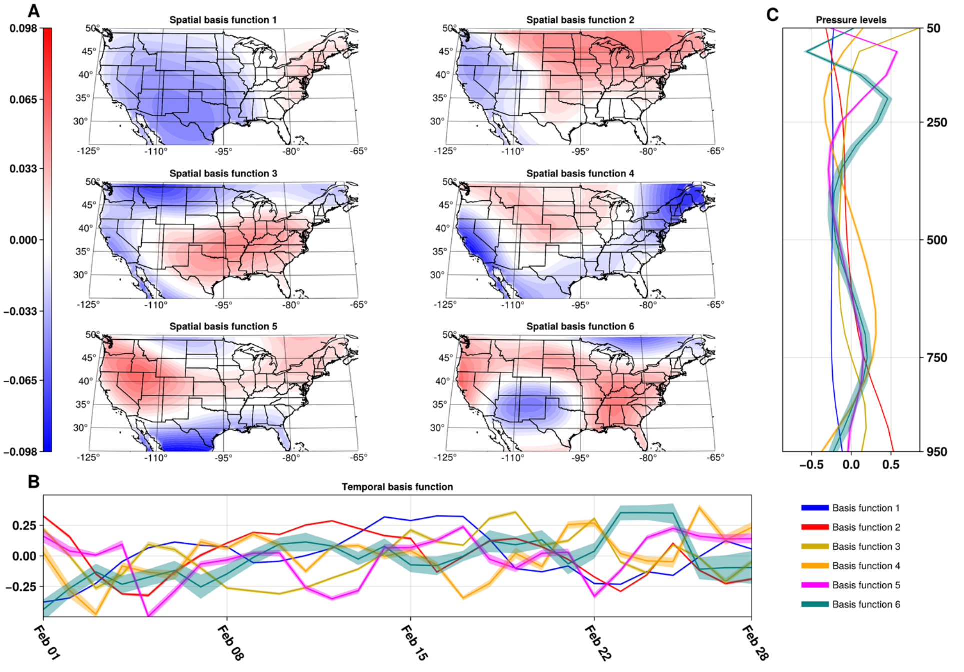

Figure 3. Atmospheric pressure basis functions: (A) Posterior mean of the first five spatial basis functions, (B) posterior mean (solid line) and 95% credible intervals (shaded region) for the first five temporal basis functions, and (C) posterior mean (solid line) and 95% credible intervals (shaded region) for the first five pressure-level basis functions .

sponding signal strength (g) is orders of magnitude smaller. Plots containing all of the basis functions are shown in the Supplemental Material. In atmospheric science, there are spatial patterns often associated with blocking events that resemble some of the basis functions (Wang et al. 2015). In panel A, the second spatial basis function resembles a dipole pattern, which loads positively (negatively) on the eastern (western) United States. The fourth spatial basis function loads positively on the central northern portion of the United States and negatively on both coasts in a U-shaped pattern. While not necessarily physically interpretable, these patterns resemble the general blocking pattern associated with cold air outbreaks, termed an *Omega Block*, where the central United States experiences high pressure relative to the east and west coasts (Pinheiro et al. 2019). For the basis functions describing the influence of pressure levels, we see that the general trends are smoothly varying to 200mb, whereas between 200mb and 50mb the trends are more variable. There are two possible reasons for this behavior. First, the transition from the troposphere to the stratosphere occurs between 100-200mb, and these two layers of the atmosphere are weakly coupled and not well mixed. Second, because air density exponentially decreases with height in the atmosphere, the distance between layers increases as the pressure level decreases (e.g., 950mb and 900mb are generally closer than 100mb and 50mb), which is not reflected in this figure.

5.2. SPECIFIC HUMIDITY: GAMMA DISTRIBUTED DATA

Humidity, the concentration of water vapor present in the air, plays a important role in weather (e.g., precipitation, dew, and fog formation; Ahrens 2014), human health (e.g., body temperature regulation; Davis et al. 2016), and terrestrial ecosystems (e.g., vapor pressure deficit; Grossiord et al. 2020), among other processes. Recently, it has been shown that climate models, such as the Coupled Model Intercomparison Project Phase 6 (CMIP6) ensemble of model simulations (Eyring et al. 2016), fail to appropriately capture humidity trends in arid regions, such as the US Southwest, compared to observations and reanalysis products (Simpson et al. 2024). Here, we show how our proposed method could be used to identify how modes of variability shape the spatial patterns of humidity, which could lead to a better understanding of the environmental drivers that shape humidity trends and how climate models may or may not be representing those drivers.

Humidity is more formally defined in atmospheric science as specific humidity, which is the ratio of the mass of water vapor to the total mass of air. We take monthly-averaged specific humidity from ERA5 (Hersbach et al. 2020) between January 2015 through December 2024 at every 50 mb from 950mb to 500mb. We regrid the ERA5 output to 1°, and focus only on the locations over land contained within 125°W to 95°W and 25°N to 50°N (see Figure S.11 in the Supplemental Material). Here, we limit the upper pressure level to 500mb because there is very low specific humidity above that level, resulting in a total of 751,200 data points. To account for an influx of water vapor into the region of interest, we obtain monthly-averaged evaporation (EV) at the surface level (1000 mb) from ERA5 over the same temporal and spatial domain, which exhibits a strong annual cycle. We also wish to account for a semi-annual cycle (SAC), parameterized by a Fourier expansion, and the effect of elevation. Initial empirical analysis suggests the influence of EV, SAC, and elevation all vary spatially. To account for this variation, we employ a spatial basis expansion of the regression coefficients using Wendland basis functions. Additionally, the amount of water vapor in the air is expected to decrease with altitude, based on the fact that air density decreases exponentially with height in the atmosphere, so we allow each pressure level to have a unique intercept. Initial empirical analysis indicates specific humidity can be well approximated by a Gamma distribution, and we proceed under this assumption, modeling the data using the R-Gamma distribution. For the random effect, we choose a rank $R = 5$ and specify a Matérn kernel with smoothness equal to 3.5 for the spatial basis functions, and a Gaussian kernel for the temporal and pressure level basis functions. We chose $R = 5$ based on the first 5 basis functions of the de-meaned log specific humidity explain approximately 90% of the variation in the data as determined from the CP-decomposition. For all three kernels, the length-scales are estimated. All other specific model details are deferred to the Supplemental Material.

Figure 4 shows estimates of the first three basis functions over (A) space, (B) time, and (C) pressure level. Similar to the previous example, we only show the first three basis functions because the remaining basis functions explain little variation in the humidity random effect. Plots containing all of the basis functions are shown in the Supplemental Material. The first basis function loads negatively over most of the region, the second basis function distinguishes the southwest from the northeast areas of the region, and the third

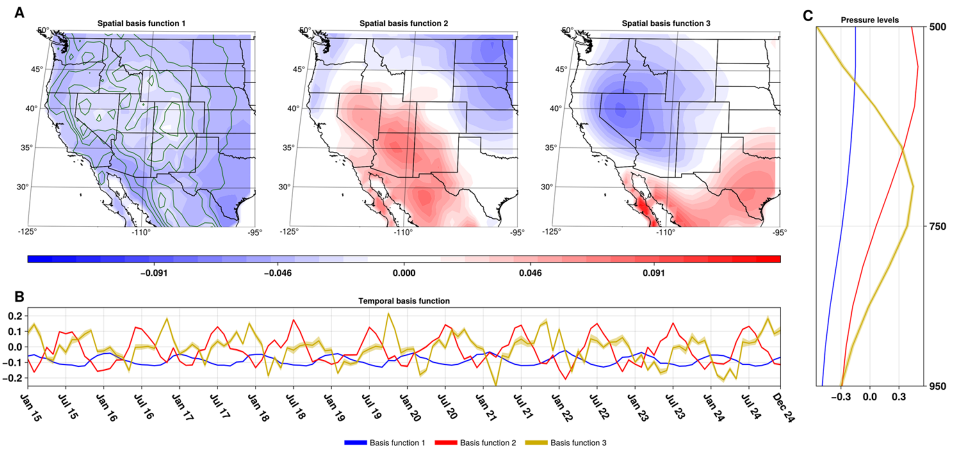

Figure 4. The Specific humidity basis functions: (A) Posterior mean of the first three spatial basis functions, (B) posterior mean (solid line) and 95% credible intervals (shaded region) for the first three temporal basis functions, and (C) posterior mean (solid line) and 95% credible intervals (shaded region) for the first three pressure-level basis functions. For the first panel of (A), the contour lines depict elevation contours .

basis function distinguishes the south-southwest from the northeast regions. While the spatial basis functions are not necessarily physically interpretable, the first spatial basis function corresponds well to the overlain elevation contours and the second visually corresponds to the North American monsoon cycle (Duan et al. 2024). In addition to the temporal patterns, the influence of the spatial basis functions changes with pressure level and is less variable than the temporal basis functions, as indicated by the CI.

Figure S.12 shows the posterior means of the spatially-varying parameter estimates. Here, we elect to show the amplitude and phase of the Fourier coefficients instead of the original parameter values, as the amplitude and phase have a physical relationship. To compute the amplitude and phase, let $a(s)$ be the spatially varying parameter associated with semi-annual cosine term in $\mathbf{X}_F$ and $b(s)$ the spatially varying parameter associated with the semi-annual sine term in $\mathbf{X}_F$. The semi-annual amplitude is $A(s) = \sqrt{a(s)^2 + b(s)^2}$ and the phase is $\varphi(s) = \text{atan2}(b(s), a(s))$. We see there is spatial variation in all covariates. Interestingly, we see elevation does not have a strong signal, which is likely due to the pressure level specific intercept accounting for specific humidity's relationship to pressure. Lastly, we see evaporation is heavily influential for the interior of the United States, but not so much for the west coast.

6. DISCUSSION

With the ever-increasing size and complexity of Earth system data, there is a pressing need for statistical methods able to extract patterns from dependent, multi-dimensional, non-stationary, and non-Gaussian data. In this paper, we present a flexible generalized linear mixed effects model for tensor-valued data that provides one possible approach to address this problem. Importantly, the model allows us to estimate the individual components of the non-stationary and non-separable random effect, parameterized using the CP-decomposition. The orthonormal matrices that comprise the CP-decomposition often have a dependent structure that is important to the user, and our methodology can identify these structures for high-dimensional problems. Additionally, we discussed how to incorporate a variety of fixed effects, such as scalar, vector, matrix, and tensor-indexed fixed effects, and how dependence can be incorporated within the fixed effects. We illustrated the method's capabilities using two different distributional assumptions, the Gaussian and Gamma distributions. Lastly, we applied our method to two real-world processes: inferring the anomalous structure of atmospheric pressure during a major blocking event and identifying the spatial patterns of some drivers of humidity over a decadal period. Additionally, while both of our applications used gridded data, the method can be applicable to non-uniformly spaced data, so long as no data is missing. While our focus is on tensor-valued data from the Earth system sciences, the framework is generally applicable to other scientific domains.

All of the examples presented here are concerned with inference for a single high-dimensional process (e.g., pressure *or* humidity). Often, interest is in performing joint inference for multiple systems, for example, inference for pressure *and* humidity or between different climate model outputs (Schliep and Hoeting 2013; Genton and Kleiber 2015). Within the proposed framework, this could be achieved by mapping each observed process to a common latent process. For example, and in keeping with the notation in Section 3 and assuming each process is a matrix-valued for notational ease, let

$$\begin{aligned}
\mathbf{Z}_1 &\sim f_1(g_1(\mathbf{H}_1|\boldsymbol{\phi}_1, \boldsymbol{\tau}_1)) \\
\mathbf{Z}_2 &\sim f_2(g_2(\mathbf{H}_2|\boldsymbol{\phi}_2, \boldsymbol{\tau}_2)) \\
&\vdots \\
\mathbf{Z}_N &\sim f_N(g_N(\mathbf{H}_N|\boldsymbol{\phi}_N, \boldsymbol{\tau}_N)),
\end{aligned}$$

where $f_i(\cdot)$ and $g_i(\cdot)$, $i = 1, \ldots, N$, denote the distribution and link function for each process, respectively. Within the latent process, the kernel function parameterizing the column correlation for the orthonormal matrix representing the process relationship should be the identity kernel (e.g., $\boldsymbol{\Omega}^{(N)} \equiv \mathbf{I}$), unless some notion of distance between the processes can be defined along with a valid kernel function. Note, this does not imply independence between processes, but instead implies there is no structure *within* the columns of the orthonormal matrix representing the process relationship. However, a limitation of our approach is the requirement that each matrix within the CP-decomposition needs to have the same rank R. This would be potentially limiting if only two systems were considered, because a rank two matrix for a high-dimensional process would likely be an unrealistic assumption. In a similar fashion to the multi-system extension, missing data can be accounted for by including a mapping function from a fully observed latent process $\mathcal{H}$ to the observed data $\mathcal{Z}$ (see Cressie and Wikle 2011, formethodstoaccommodatemissingdatainthespatio-temporalsetting). We do not explore these components here, but future work could investigate these additional features and the modeling implications.

Our framework is general in that it is defined for N-dimensional tensors. However, practical application is still limited by computational resources. That is, because inference is performed jointly for model components, all data is processed simultaneously. For example, because we are concerned with data $\mathcal{Z} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, if $I_k = 10$ for $k = 1, \ldots, 10$, this results in 10 billion data points (roughly 80 GB) which could be prohibitive to load into memory for many personal computers. A very active area of research for computer science is on how best to perform operations on massive data, and one direction of future work could be directed at incorporating these methods of handling massive data into the statistical work presented here. Related is the computational cost of our approach. While we can reduce the computational cost of estimating dependent data from $\mathcal{O}((I_1 \times \cdots \times I_N)^3)$ to $\mathcal{O}(max\{I_1^3, \cdots, I_N^3\})$, our approach is still dependent on inverting matrices for each dimension. Depending on the size of the dimensions, this can prohibit the use of our approach. Within the Gaussian Process literature, there is significant interest in methods reducing this computational bottleneck, such as sparse kernels and approximate methods (see Liu et al. 2019, forarecentreview). Future work could focus on leveraging these advancements within our proposed methodology.

ACKNOWLEDGEMENTS

This research was supported by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 and by the Regional and Global Model Analysis Program area within the Earth and Environmental Systems Modeling Program as part of the Calibrated and Systematic Characterization, Attribution, and Detection of Extremes (CASCADE) project. The authors would like to thank Micahel F. Wehner for help providing a physical explanation to the modes of variability for the real-world applications in addition to other insight about the methods and application areas of interest. The authors would also like to thank Mariah North for help with visualizations, especially in choosing color palettes.

Data Availability All data and code are publically available at https://github.com/jsnowynorth/BayesianHOSVD.jl. Data that are not included in the directory are publically downloadable via the python script included in the *python* folder.

Declarations
Conflicts of Interest The authors declare no conflicts of interest.

[Received June 2025. Revised May 2026. Accepted May 2026.]

REFERENCES

Ahrens CD (2014) Essentials of Meteorology: An Invitation to the Atmosphere. Cengage Learning.(7 edition)

Albert JH, Chib S (1993) Bayesian analysis of binary and polychotomous response data. J Am Stat Assoc 88(422):669–679

Banerjee S, Carlin BP, Gelfand AE (2015) Hierarchical modeling and analysis for spatial data. Chapman and Hall/CRC, Bezanson J, Edelman A, Karpinski S, Shah VB (2017) Julia: A fresh approach to numerical computing. SIAM Rev 59(1):65–98

Billio M, Casarin R, Iacopini M, Kaufmann S (2023) Bayesian dynamic tensor regression. Journal of Business & Economic Statistics 41(2):429–439

Byrne S, Girolami M (2013) Geodesic Monte Carlo on embedded manifolds. Scand J Stat 40(4):825–845

Carroll JD, Chang J-J (1970) Analysis of individual differences in multidimensional scaling via an n-way generalization of â€œEckart-Young-decomposition. Psychometrika 35(3):283–319

Chatzichristos C, Kofidis E, Morante M, Theodoridis S (2019) Blind fMRI source unmixing via higher-order tensor decompositions. J Neurosci Methods 315:17–47

Chikuse Y (2003) Statistics on Special Manifolds, vol 174. Lecture Notes in Statistics. Springer, New York, New York, NY

Cressie NAC, Wikle CK (2011) Statistics For Spatio-Temporal Data. John Wiley & Sons

Davis NA, Richter JH, Glanville AA, Edwards J, LaJoie E (2022) Limited surface impacts of the January 2021 sudden stratospheric warming. Nat Commun 13(1):1136

Davis RE, McGregor GR, Enfield KB (2016) Humidity: A review and primer on atmospheric moisture and human health. Environ Res 144(January):106–116

De Lathauwer L, De Moor B, Vandewalle J (2000) A multilinear singular value decomposition. SIAM J Matrix Anal Appl 21(4):1253–1278

Duan S, Ullrich P, Boos WR (2024) Meteorological drivers of North American monsoon extreme precipitation events. J Geophys Res (13):129Atmospheres

Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE (2016) Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geoscientific Model Development 9(5):1937–1958

Finley AO, Banerjee S, Ek AR, McRoberts RE (2008) Bayesian multivariate process modeling for prediction of forest attributes. J Agric Biol Environ Stat 13(1):60–83

Fuentes M (2001) A high frequency kriging approach for non-tationary environmental processes. Environmetrics 12(5):469–483

Fuentes M, Chen L, Davis JM (2008) A class of nonseparable and nonstationary spatial temporal covariance functions. Environmetrics 19(5):487–507

Gelfand AE, Schmidt AM, Banerjee S, Sirmans CF (2004) Nonstationary multivariate process modeling through spatially varying coregionalization. TEST 13(2):263–312

Genton MG, Kleiber W (2015) Cross-covariance functions for multivariate geostatistics. Stat Sci 30(2):147–163

Glanz H, Carvalho L (2018) An expectation–maximization algorithm for the matrix normal distribution with an application in remote sensing. J Multivar Anal 167:31–48

Grossiord C, Buckley TN, Cernusak LA, Novick KA, Poulter B, Siegwolf RT, Sperry JS, McDowell NG (2020) Plant responses to rising vapor pressure deficit. New Phytol 226(6):1550–1566

Guhaniyogi R, Qamar S, Dunson DB (2015) Bayesian tensor regression. J Mach Learn Res 18:1–31

Guhaniyogi R, Spencer D (2021) Bayesian tensor response regression with an application to brain activation studies. Bayesian Anal 16(4):1221–1249

Hannachi A, Jolliffe IT, Stephenson DB (2007) Empirical orthogonal functions and related techniques in atmospheric science: A review. Int J Climatol 27(9):1119–1152

Harman HH, Harman HH (1976) Modern factor analysis. University of Chicago Press

Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, Simmons A, Soci C, Abdalla S, Abellan X, Balsamo G, Bechtold P, Biavati G, Bidlot J, Bonavita M, Chiara G, Dahlgren P, Dee D, Diamantakis M, Dragani R, Flemming J, Forbes R, Fuentes M, Geer A, Haimberger L, Healy S, Hogan RJ, Hólm E, Janisková M, Keeley S, Laloyaux P, Lopez P, Lupu C, Radnoti G, Rosnay P, Rozum I, Vamborg F, Villaume S, Thépaut J (2020) The ERA5 global reanalysis. Q J R Meteorol Soc 146(730):1999–2049

Hoff PD (2007) Model averaging and dimension selection for the singular value decomposition. J Am Stat Assoc 102(478):674–685

Hoff PD (2009) Simulation of the matrix Bingham–von Mises–Fisher distribution, with applications to multivariate and relational data. J Comput Graph Stat 18(2):438–456

Hoff PD (2011) Separable covariance arrays via the Tucker product, with applications to multivariate relational data. Bayesian Anal 6(2):179–196

Huang A, Wand MP (2013) Simple marginally noninformative prior distributions for covariance matrices. Bayesian Anal 8(2):439–452

Hughes J, Haran M (2013) Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. J R Stat Soc Ser B Stat Methodol 75(1):139–159

Jauch M, Düker M-C, Hoff P (2025) Prior distributions for structured semi-orthogonal matrices. ArXiv 1

Jauch M, Hoff PD, Dunson DB (2021) Monte Carlo simulation on the Stiefel manifold via polar expansion. J Comput Graph Stat 30(3):622–631

Kang EL, Cressie N (2011) Bayesian inference for the spatial random effects model. J Am Stat Assoc 106(495):972–983

Kapteyn A, Neudecker H, Wansbeek T (1986) An approach ton-mode components analysis. Psychometrika 51(2):269–275

Katzfuss M, Guinness J (2021) A general framework for Vecchia approximations of Gaussian Processes. Stat Sci 36(1):124–141

Kiers HAL (2000) Towards a standardized notation and terminology in multiway analysis. J Chemom 14(3):105–122

Kolda TG (2006) Multilinear operators for higher-order decompositions. Sandia National Laboratories (SNL). Albuquerque, NM, and Livermore, CA (United States,(Technical Report)

Kolda TG, Bader BW (2009) Tensor decompositions and applications. SIAM Rev 51(3):455–500

Korevaar H, Metcalf CJ, Grenfell BT (2020) Tensor decomposition for infectious disease incidence data. Methods Ecol Evol 11(12):1690–1700

Lei M, Labbe A, Sun L (2024) Scalable spatiotemporally varying coefficient modeling with Bayesian kernelized tensor regression. *Bayesian Analysis*, -1(-1):1–29

Liu H, Ong Y-S, Shen X, Cai J (2019) When Gaussian Process meets big data: A review of scalable GPs. IEEE Transactions on Neural Networks and Learning Systems 31(11):4405–4423

Lock EF (2018) Tensor-on-tensor regression. J Comput Graph Stat 27(3):638–647

Lorenz EN (1956) Empirical orthogonal functions and statistical weather prediction, vol 1. Massachusetts Institute of Technology, Department of Meteorology Cambridge

Mantua NJ, Hare SR (2002) The Pacific decadal oscillation

Mardia KV, Jupp PE (1999) Directional Statistics. Wiley Series in Probability and Statistics. John Wiley & Sons Inc, Hoboken, NJ, USA

Neal R (2011) MCMC Using Hamiltonian Dynamics. In *Handbook of Markov Chain Monte Carlo*, pages 113–162. CRC press

North GR, Bell TL, Cahalan RF, Moeng FJ (1982) Sampling errors in the estimation of empirical orthogonal functions. Mon Weather Rev 110(7):699–706

North JS, Risser MD, Breidt FJ (2024) A flexible class of priors for orthonormal matrices with basis function-specific structure. Spatial Statistics 64:100866

O'Brien JP, Deser C (2023) Quantifying and understanding forced changes to unforced modes of atmospheric circulation variability over the North Pacific in a coupled model large ensemble. J Clim 36(1):19–37

Pinheiro MC, Ullrich PA, Grotjahn R (2019) Atmospheric blocking and intercomparison of objective detection methods: Flow field characteristics. Clim Dyn 53(7–8):4189–4216

Porcu E, Furrer R, Nychka D (2021) 30 Years of space–time covariance functions. Wiley Interdisciplinary Reviews Computational Statistics 13(2):1–24

Pourzanjani AA, Jiang RM, Mitchell B, Atzberger PJ, Petzold LR (2021) Bayesian enference over the Stiefel manifold via the Givens representation. Bayesian Anal 16(2):639–666

Schliep EM, Hoeting JA (2013) Multilevel latent Gaussian process model for mixed discrete and continuous multivariate response data. J Agric Biol Environ Stat 18(4):492–513

Shivadekar S, Yesha Y, Halem M, Yang Z (2023) Multi-variable tensor decomposition analytics. In *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, volume 2023-July, pages 3885–3888. IEEE

Sidiropoulos ND, De Lathauwer L, Fu X, Huang K, Papalexakis EE, Faloutsos C (2017) Tensor decomposition for signal processing and machine learning. IEEE Trans Signal Process 65(13):3551–3582

Simpson IR, McKinnon KA, Kennedy D, Lawrence DM, Lehner F, Seager R (2024) Observed humidity trends in dry regions contradict climate models. Proc Natl Acad Sci USA 121(1):1–12

Thompson DWJ, Wallace JM (2000) Annular modes in the extratropical circulation. Part I: Month-to-month variability*. J Clim 13(5):1000–1016

Tucker LR (1966) Some mathematical notes on three-mode factor analysis. Psychometrika 31(3):279–311

Vasilescu MAO, Terzopoulos D (2002) Multilinear analysis of image ensembles: TensorFaces. In *Computer Vision — ECCV 2002*, volume 2350, pages 447–460

Vecchia AV (1988) Estimation and model identification for continuous spatial processes. J R Stat Soc Ser B Stat Methodol 50(2):297–312

Wang K, Xu Y (2024) Bayesian tensor-on-tensor regression with efficient computation. Statistics and Its Interface 17(2):199–217

Wang SS, Huang W, Yoon J (2015) The North American winter dipole and extremes activity: a CMIP5 assessment. Atmospheric Science Letters 16(3):338–345

Wikle CK, Zammit-Mangion ACN (2019) Spatio-Temporal Statistics with R. Chapman and Hall/CRC. CRC Press,

Zhou H, Li L, Zhu H (2013) Tensor regression with applications in neuroimaging data analysis. J Am Stat Assoc 108(502):540–552