METHODS FOR COMPARISON AND ANALYSIS OF SPATIOTEMPORAL FIELDS

BY

ROBERT CHARLES GARRETT

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Statistics
in the Graduate College of the
University of Illinois Urbana-Champaign, 2024

Urbana, Illinois

Doctoral Committee:

Professor Bo Li, Chair
Professor Jeffrey A. Douglas
Adjunct Clinical Assistant Professor Lyndsay Shand
Assistant Professor Trevor A. Harris, Texas A&M University

Abstract

This dissertation develops three methods for spatiotemporal fields data, each designed to address research topics in climate science. The first two methods are similarity measures for evaluating the differences between climate models and observational datasets. The last method is a multivariate spatiotemporal model which characterizes the joint evolution of observed climate processes.

Chapter 2 introduces the spherical convolutional Wasserstein distance (SCWD) to more comprehensively measure differences between climate models and observational data. This new similarity measure accounts for spatial variability using convolutional projections and quantifies local differences in the distribution of climate variables. We apply SCWD to evaluate the historical model outputs of the Coupled Model Intercomparison Project (CMIP) members by comparing them to observational and reanalysis data products. Additionally, we investigate the progression from CMIP phase 5 to phase 6 and find modest improvements in the phase 6 models regarding their ability to produce realistic climatologies.

Chapter 3 introduces the sliced elastic distance, a new metric which considers potential time misalignment between climate models and observational data. Sliced elastic distance decomposes differences in the local evolution of climate processes into shape differences (amplitude), timing variability (phase), and bias (translation). We apply the sliced elastic distance to rank CMIP phase 6 precipitation models by their similarity to observational data at both global and regional scales. Using intermediate calculations from our method, we perform an in-depth phase analysis of the Indian summer monsoon to identify timing biases in the onset and retreat of the monsoon season in each CMIP6 model.

Finally, Chapter 4 introduces a novel multivariate space-time dynamic model to quantify relationships in the joint evolution of atmospheric processes. This model captures spatial variation using a flexible set of basis functions for which the coefficients are allowed to vary in time through a vector autoregressive (VAR) structure. The model is cast in a Bayesian dynamic linear model (DLM) framework and estimated using a customized MCMC sampling approach. We apply this model to study the relationship between aerosols, radiation, and temperature following the 1991 Mt. Pinatubo eruption and highlight when such a model is advantageous over simpler univariate models.

For my wife, Katherine, who stood by me every step of this journey. I love you.

Acknowledgments

I would like to thank all my teachers over the years who made this possible. In particular, I want to thank my advisor, Bo Li, who has supported my career and research in many ways including connecting me with fellowship and internship opportunities, and whose time, effort, guidance, and advice was invaluable in completing my research the last five years.

I would also like to thank my other committee members Trevor Harris, Lyndsay Shand, and Jeff Douglas, as well as collaborators Gabriel Huerta and Zhuo Wang, for their efforts and insights. I would like to express additional gratitude to Trevor and Lyndsay for their mentorship, advice, and many interesting discussions outside of regular meeting times. Also, thank you to Thomas Fisher at Miami University, who motivated me to pursue a career in research. Thank you to the National Science Foundation and DIGI-MAT team for supporting my research and education, and Sandia National Laboratories and the CLDERA Grand Challenge team for supporting my career and providing many resources and opportunities.

A special thanks to all of my family, including my parents, Pam and Chris, who inspired me to study statistics and who have always supported me, and my wife, Katherine, for her love, patience, and unwavering support through her own graduate degree. Thank you to Bob Krueger (my best man) for being an amazing friend and for playing the role of a statistician every once in a while. Finally, thank you to so many other friends I've made during this experience for all the great times we've had.

Table of contents

Chapter 1

Introduction

Anthropogenic climate change presents several concerns for society, including impacts on the environment, agriculture, and sea levels as well as increases in extreme events such as droughts, wildfires, and floods. Climate science, which is focused on understanding interactions between the Earth's climate processes, is therefore essential for understanding the sources of these changes as well as modeling their impacts. To this end, climate scientists have developed observational data products, which capture a historical record of global weather conditions, as well as climate models (also called General Circulation Models) which are computer simulations of the physical processes that drive the climate system. Observational datasets provide a baseline source of information on climate conditions and dynamics, allowing for the study and detection of trends such as global warming. These datasets also serve as the basis for evaluating climate models, which are essential for understanding future climate changes in response to factors such as increased CO_2 emissions. Both data sources are often recorded for many locations and time periods.

Comparing and analyzing climate model outputs and observational datasets is challenging due to the high-dimensional nature of such data as well as the spatially- and temporally-correlated dependency structure. Many traditional statistical methods are inadequate to account for this data structure. So, new statistical approaches have been developed to study climate data, including recent methods from the fields of spatial statistics and functional data analysis.

Spatial statistics features a wide range of techniques that are focused on accounting for and understanding the spatial dependency structure of data observed over different locations. Methods in spatial statistics may also capture trends for spatial data that vary over time or contain multiple variables. These capabilities are particularly useful for understanding the dynamics (temporal evolution) and interactions (multivariate dependencies) of spatiotemporal fields data, making spatial statistics a popular choice for prediction and

inference tasks in climate science.

Similar to spatial statistics, functional data analysis is another popular field for working with data that feature complex dependency structures. In functional data analysis, the basic data unit is assumed to be a continuous function on a domain of interest. This domain traditionally represents the time dimension, however other options have been considered. For climate data, which we consider as spatiotemporal fields, we can additionally incorporate the spatial region of interest into the functional structure. The climate data considered in this work are observed globally, so we can specify a spherical function domain to account for the geometry of the Earth's surface. Representing climate data as functions of space or functions of both space and time is useful for climate model evaluation because it provides a natural framework for comparing features such as the location and timing of events between two data products.

Chapters 2 and 3 focus on developing metrics for spatiotemporal fields comparison and offer two different perspectives on climate model evaluation. Climate model evaluation methods typically rely on direct comparisons of the long-term mean climate state at a local or global scale. However, many important climate features are not accounted for in such analyses. For example, interannual variations in the climate, which include extreme events such as heat/cold waves, droughts, and floods, are not well-characterized when taking averages over many years. To assess differences in the full climate distribution of climate models and observational data, Chapter 2 introduces a new similarity measure called the spherical convolutional Wasserstein distance (SCWD). This method views climate models and observational data as distributions of functions on the sphere, each of which represents climate conditions for a single day across the Earth's surface. Inspired by recent "sliced Wasserstein" techniques in the machine learning literature, SCWD uses a novel functional convolution slicing technique that isolates one-dimensional spatial features of these high-dimensional climate data to enable efficient computations. The resulting distance provides a global summary of local distributional deviations for each climate model, and the intermediate calculations can be used to produce spatial maps which highlight the regions of greatest variability. We apply SCWD to compare climate models for both temperature and precipitation to historical datasets, and find overall lower distances for the Coupled Model Intercomparison Project phase 6 (CMIP6) models compared to the previous phase 5 (CMIP5) models, marking an improvement in the ability of climate models to accurately reconstruct local distributions.

Another climate feature that is not accounted for in most traditional climate model evaluation approaches is phase variability, or deviations in the timing of climate events. Timing is of keen interest when evaluating the performance of models in regions with extreme seasonal variability, including monsoon regions. To quantify phase variability in climate models, Chapter 3 introduces another new metric called the sliced elastic distance. This new metric extends the ordinary elastic distance, which features a phase alignment step that is valid for functions of time only, to compare functions with a joint space-time domain using a convolutional

slicing approach similar to SCWD. More generally, we show that this slicing approach extends arbitrary metrics and vector-valued metrics for functions of time to create new metrics for functions of space and time. We apply sliced elastic distance to rank precipitation models by their agreement with historical observations and identify a cohort of models for further study. Then, using intermediate results from the phase alignment step, we perform an in-depth analysis of this cohort in the Indian summer monsoon region to quantify local biases in the timing of the onset and retreat of the monsoon. We find that some models feature little to no timing bias in the monsoon arrival, while others exhibit early or late arrival biases. For the monsoon retreat, we find that almost all models in our cohort feature late timing biases compared to the observations, indicating that the usual method of comparing the average precipitation over the typical observed monsoon period may be insufficient when evaluating climate model performance in this region.

Chapter 4 focuses on a different topic in climate science: analyzing the interactions between atmospheric processes. This work is motivated by the June 1991 Mt. Pinatubo eruption, which released a massive amount of volcanic aerosols into the atmosphere. Because of the huge magnitude of this event, it serves as a natural analog for stratospheric aerosol injection, which is a proposed approach to modify solar radiation and thus counteract the effects of global warming. The effects of this event covered many different regions, time periods, and climate variables, making it challenging to analyze with previous approaches in spatial statistics. So, we developed a novel multivariate space-time dynamic model (MV-STDM) to quantify and characterize these multifaceted impacts. Spatial variability in our MV-STDM is modeled through a basis function approach which provides a low-rank representation of atmospheric processes through a state vector of basis function coefficients. The multivariate and temporal evolution of the processes is captured through a transition matrix which follows a vector autoregressive (VAR) structure. We estimate the model through a customized Markov Chain Monte Carlo (MCMC) approach, and demonstrate the effectiveness of this approach on a synthetic dataset. Finally, we apply our MV-STDM to model aerosol optical depth, outgoing longwave radiation, and stratospheric temperature anomalies following the Mt. Pinatubo eruption and find evidence of spatially-varying trends in the evolution of each atmospheric process. We also demonstrate the performance of our MV-STDM in predicting stratospheric temperatures over North America in the years following the eruption, highlighting our model's advantages over simpler approaches.

In summary, this dissertation introduces three new methods to study important topics in climate science. A key focus shared by all three methods is interpretability: each metric or model offers insights beyond just a one-number summary, promoting transparency and providing useful information for climate scientists. The contributions include comprehensive tools for evaluating climate models and for understanding the evolution of climate processes, and the findings highlight the flexibility and utility of spatial statistics and functional data analysis for addressing challenges in climate science.

Chapter 2

Validating Climate Models with Spherical Convolutional Wasserstein Distance

2.1 Introduction

[1]General Circulation Models, or climate models, are mathematical representations of the climate system that describe interactions between matter and energy through the ocean, atmosphere, and land [Washington and Parkinson, 2005]. Climate models are the primary tool for investigating the response of the climate system to changes in forcing, such as increases in CO_2, and projecting future climate states [Flato et al., 2014]. To assess the plausibility of climate models, climate scientists compare output from model simulations against observational data [Rood, 2019]. This comparison is the focus of climate model validation techniques for ensuring that climate models capture the dynamics of the climate system [Roca et al., 2021].

The Coupled Model Intercomparison Project (CMIP) was initiated in 1995 as a comprehensive and systematic program for assessing climate models against each other and observational data [Eyring et al., 2016a]. Each model in CMIP participates in a wide variety of experiments such as performing a historical simulation, a pre-industrial control simulation, and various simulations representing different scenarios for CO_2 emissions [Eyring et al., 2016a]. Because historical simulations coincide with observational measurements, we can compare each model's synthetic climate distribution to the distribution of observational or quasi-

[1]This chapter is based on the following manuscript in review: Garrett, R. C., Harris, T., Li, B., & Wang, Z. (2024). Validating Climate Models with Spherical Convolutional Wasserstein Distance. arXiv preprint arXiv:2401.14657.

observational data products [Raäisaänen, 2007], to assess their reconstructive skill. For complete spatial coverage we compare against reanalysis data, a blend of observations and short-range weather forecasts through data assimilation [Bengtsson et al., 2004]. This has become one popular climate model validation method [Flato et al., 2014].

Previous approaches Many statistical and machine learning-based methods have been applied to assess climate model output against reanalysis fields. The most common approach is to compute the root mean square error (RMSE) between long-term means of the climate model output and the reanalysis field [Li et al., 2021, Zamani et al., 2020, Karim et al., 2020, Ayugi et al., 2021]. RMSE provides a direct measure of the differences between two climate fields but does not take into account internal variability, so we should use it with caution in evaluating climate models. Another approach, which is invariant to bias, is to compute measures of correlation between climate model output and reanalysis fields [Zhao et al., 2021, Zamani et al., 2020, Karim et al., 2020, Ayugi et al., 2021].

More comprehensive approaches employ techniques for random processes to compare two spatial fields. For example, Shen et al. [2002] and Cressie et al. [2008] use the wavelet decomposition to compare the spatial-frequency content of two fields. Hering and Genton [2011] measures the loss differential between models of spatial processes, Lund and Li [2009] and Li and Smerdon [2012] compare the first and second moments of two random processes, and Yun et al. [2022] identifies local differences in the mean and dependency structure between two spatiotemporal climate fields. Functional data analysis techniques have been introduced to compare spatial and spatiotemporal random fields, considering the random fields as continuous functions. Many of these approaches compare the underlying mean functions from two sets of functional data [Zhang and Chen, 2007, Horváth et al., 2013, Staicu et al., 2014]. Other approaches include the second-order structure in the comparison [Zhang and Shao, 2015, Li et al., 2016], or compare the distributions of two spatial random processes [Harris et al., 2021a].

Since climate models aim to mimic the real climate which is the underlying pattern of weather, directly assessing the distributional differences between the modeled and observed data seems a more thorough approach to evaluating climate models. Vissio et al. [2020] proposed to use the Wasserstein distance (WD), a popular metric for comparing probability distributions [Villani, 2009], for a such purpose. The WD can be computationally expensive or even impossible to calculate between high-dimensional distributions [Kolouri et al., 2019], so Vissio et al. [2020] first converts each climate field to a single spatial mean and then only compares the distribution of spatial means. However, this dimension reduction puts their method at the risk of missing important spatial variability information, and consequently failing to accurately distinguish two climate fields that are different.

Recent contributions in the Machine Learning literature seek to compare multivariate distributions using many features while leveraging the efficiency of the one-dimensional WD [Vallender, 1974]. The sliced WD [Bonneel et al., 2015] compares random projections of distributions on $\mathbb{R}^n$, and the generalized sliced WD [Kolouri et al., 2019] extends this to a broader class of projections. The convolutional sliced WD [Nguyen and Ho, 2022] compares distributions of discrete square images in the space $\mathbb{R}^{n\times n}$, accounting for the spatial structure using kernel convolutions. However, spatial fields from climate models are defined on a spherical domain, making them incompatible with the vectorized/square nature of these distances. The spherical sliced WD [Bonet et al., 2022] and other spherical transport methods [Quellmalz et al., 2023, Cui et al., 2019] can compare distributions of spatial point processes over a sphere, such as locations of natural disasters and extreme weather events, but cannot be used for smooth fields such as daily temperature and precipitation.

Our proposal We propose the functional sliced WD as a generalization of the sliced WD to distributions of functional data. To create a tailored tool for climate model evaluation, we define the Spherical Convolutional Wasserstein distance (SCWD) as a special case of the functional sliced WD for functions on the unit sphere $\mathbb{S}^2$, a manifold on which latitude-longitude coordinates are defined. SCWD creates slices containing a small region of spatial fields to characterize local differences in the distribution of climate variables, which are further integrated into a single measure for global differences. Compared to the spatial mean-based WD from Vissio et al. [2020], SCWD accounts for spatial features while maintaining the modest computation for univariate WD when comparing climate models against reanalysis data, resulting in a comprehensive and more robust evaluation. We apply SCWD to rank climate models and assess the progression of the new CMIP era with respect to daily average surface temperature and daily total precipitation.

2.2 Preliminaries

We consider the problem of comparing two probability distributions P and Q. Each distribution is a member of $\mathcal{P}(\Omega)$, the set of Borel probability measures on some sample space, Ω. In our application, we treat climate fields as functional data [Wang et al., 2016] over a spatial domain $\mathcal{S}$, thus we often assume Ω is a space of functions. Our sample space of interest is $L^2(\mathcal{S})$, the set of square integrable functions from $\mathcal{S} \to \mathbb{R}$ where $\mathcal{S}$ is a compact subset of $\mathbb{R}^n$. CMIP model outputs are available at a global scale, so we consider the spatial domain to be the unit sphere $\mathbb{S}^2$, the space over which latitude and longitude coordinates are assigned to the Earth's surface. In fact, the space $L^2(\mathbb{S}^2)$ has been previously considered for modeling climate fields [Heaton et al., 2014]. To compare two functional data distributions P and Q, we define a distance function $D(P, Q)$ to act as a similarity measure.

Comparing Distributions of Functions Comparisons for distributions of functions in $L^2(\mathcal{S})$ have been studied for specific cases of $\mathcal{S}$ [Hall and Van Keilegom, 2007, Bugni and Horowitz, 2021, Pomann et al., 2016, Harris et al., 2021a], but none of these have focused on $L^2(\mathbb{S}^2)$. The mathematical properties of probability measures in $\mathcal{P}(L^2(\mathcal{S}))$ have been studied before [Gijbels and Nagy, 2017, Kim, 2006], but it is challenging to calculate distances such as the WD in this space without additional assumptions [Li and Ma, 2020]. To define a similarity measure for distributions in $\mathcal{P}(L^2(\mathcal{S}))$, we build on the theory of the sliced WD and its various extensions, which have only been defined for distributions of finite-dimensional data.

Sliced WD Given a Borel function $\pi : \Omega \to \mathbb{R}$, the pushforward of P under π is a valid distribution in $\mathcal{P}(\mathbb{R})$ defined as $\pi\#P(B) = P(\pi^{-1}(B))$ for all Borel sets B in $\mathbb{R}$. The r-th order sliced WD [Bonneel et al., 2015] between $P, Q \in \mathcal{P}(\mathbb{R}^n)$ is a metric defined as the mean of the ordinary WD over univariate pushforwards:

$$SW_r(P, Q) = \left(\int_{\mathbb{S}^{n-1}} W_r(\pi_\theta \# P, \pi_\theta \# Q)^r d\theta \right)^{1/r}, \tag{2.1}$$

where $\pi_\theta(x) = x^T\theta$ for $x \in \mathbb{R}^n$ and $\theta \in \mathbb{S}^{n-1}$. We call π_θ the slicing function because it produces one-dimensional "slices" of the data using projection matrices. Because $\pi_\theta\#P$ and $\pi_\theta\#Q$ are valid distributions in $\mathcal{P}(\mathbb{R})$, the WD inside the integral can be calculated with the commonly used analytical form for univariate measures [Vallender, 1974].

Generalized Sliced WD The generalized sliced WD [Kolouri et al., 2019] replaces π_θ with a more general class of slicing functions, denoted as g_θ. Because of this increased flexibility, the generalized sliced WD is a pseudometric rather than a metric except for specific cases of g_θ [Kolouri et al., 2019]. For a given use case, the utility and metric properties of the generalized sliced WD are therefore determined by the choice of slicing function. Slicing functions can be chosen via optimization, estimated using neural networks, or specified by the researcher to isolate features of interest [Kolouri et al., 2019]. The last option is desirable for our application, allowing the slices to be restricted to spatial features of interest to climate modelers. However, the generalized sliced WD is defined between distributions of data in $\mathbb{R}^n$, a space that is not suitable for distributions of spatial fields. Climate fields could be coerced to vectors in $\mathbb{R}^n$ to be made compatible with the generalized sliced WD. However, this would result in a loss of the inherent spatial structure, making it challenging to specify a slicing function that can handle considerations such as area weighting and spatial correlations.

Convolution Sliced WD The convolution sliced WD [Nguyen and Ho, 2022] represents images as matrices in the space $\mathbb{R}^{n\times n}$. The slicing function is replaced with kernel convolutions, or possibly a sequence of kernel convolutions, of $d \times d$ pixels for $d \in \mathbb{N}$. The kernel aggregates nearby locations to isolate local features, which

could provide useful information to climate modelers. However, when climate fields are represented on a rectangular grid, the geographic area represented by each grid cell varies drastically between latitudes due to the non-Euclidean structure. Thus, a $d \times d$ pixel kernel will cover different sizes and shapes of geographic areas depending on location. For our application, the kernel radius should therefore be defined using geographic distance, not pixels.

2.3 Methods

We introduce the functional sliced WD framework which extends the flexible slicing process from the generalized sliced WD to the infinite-dimensional case of functions in $L^2(\mathcal{S})$. We focus on the special case of $L^2(\mathbb{S}^2)$, which we call the spherical convolutional WD (SCWD), for our climate model validation application. SCWD adapts the kernel convolution-based slicing idea of the convolutional sliced WD to a continuous setting while accounting for the non-euclidean structure of climate fields realized over the Earth's surface.

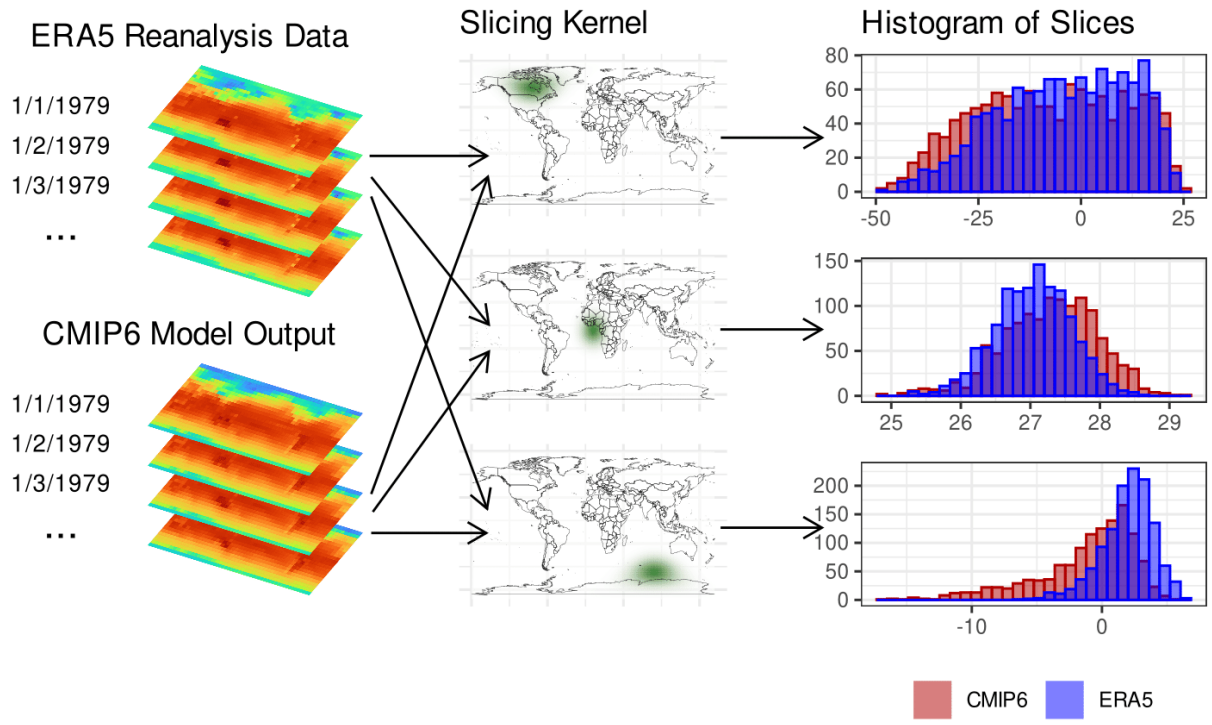


Figure 2.1: Diagram representing the calculation of SCWD between distributions of daily mean surface temperature (in degrees Celsius) from ERA5 and a CMIP6 model. Each day, many projections are computed using kernel convolutions, represented here at two locations. The resulting projections, called slices, summarize the local climate conditions in each dataset. The slices for each day are viewed as a sample from the marginal distribution at each location, represented here as histograms. SCWD is computed as a global mean over the univariate WD between each pair of local distributions.

2.3.1 Functional Sliced Wasserstein Distance

In general it is not possible to analytically characterize distributions in $\mathcal{P}(L^2(\mathcal{S}))$ and there are no closed form solutions for computing the WD. However, it is possible to slice elements of $L^2(\mathcal{S})$, meaning we can leverage the analytical form of the one-dimensional WD to define a computable sliced WD. We extend the convolution slicer from Nguyen and Ho [2022] to the functional data case, allowing us to project functions in $L^2(\mathcal{S})$ to values in $\mathbb{R}$ while preserving local spatial information.

Definition 2.1 (Convolution Slicer)**.** Let $\mathcal{S}$ be a compact subset of $\mathbb{R}^n$, $s \in \mathcal{S}$, and k, called the kernel function, be a continuous function from $\mathcal{S} \times \mathcal{S} \to [0, \infty)$. We define the convolution slicer $c_s(f)$, a linear operator from $f \in L^2(\mathcal{S}) \to \mathbb{R}$, as follows:

$$c_s(f) = \int_{\mathcal{S}} f(u)k(s,u)du.$$

Since location $s \in \mathcal{S}$ is fixed, $k(s,u)$ is a continuous function from $u \in \mathcal{S} \to \mathbb{R}$. Because $\mathcal{S}$ is compact, it follows that $k(s,u)$ is bounded and thus $k(s,u) \in L^2(\mathcal{S})$. So, by Hölder's inequality, $c_s(f)$ is a bounded linear operator from $L^2(\mathcal{S}) \to \mathbb{R}$. Bounded linear operators are also continuous [Stein and Shakarchi, 2011], so $c_s(f)$ is a continuous linear operator and, therefore, Borel measurable. Thus, for any measure $P \in \mathcal{P}(L^2(\mathcal{S}))$, the pushforward $c_s\#P$ is a valid measure in $\mathcal{P}(\mathbb{R})$. Thus, we can define a functional sliced WD between distributions in $\mathcal{P}(L^2(\mathcal{S}))$ as follows:

Definition 2.2 (Functional Sliced WD)**.** Let $\mathcal{S}$ be a compact subset of $\mathbb{R}^n$, $r \geq 1$, and $P, Q \in \mathcal{P}(L^2(\mathcal{S}))$. Let c_s be an operator satisfying Definition 2.1. We define the (r-th order) functional sliced WD between P and Q as follows:

$$FSW_r(P,Q) = \left(\int_{\mathcal{S}} W_r(c_s\#P, c_s\#Q)^r ds \right)^{1/r}$$

where W_r is the Wasserstein metric on $\mathcal{P}(\mathbb{R})$.

Because $c_s\#P$ and $c_s\#Q$ are valid univariate probability measures, the analytical form of the univariate WD can be applied for efficient calculations. We introduce theoretical properties for the functional sliced WD in Theorem 2.1.

Theorem 2.1. *For all compact subsets $\mathcal{S} \subset \mathbb{R}^n$, FSW_r is a pseudometric on $\mathcal{P}(\mathcal{F}_{\mathcal{S}})$ and maintains the r-convexity property of the ordinary W_r metric.*

Proof of Theorem 2.1 is provided in Section 2.7.1. It is unknown if the final positivity property of a metric is satisfied. Proof of this property would require an invertible Radon-like transformation to be defined for probability measures in $\mathcal{P}(L^2(\mathcal{S}))$. Therefore, as with the generalized sliced WD, it is up to the researcher

to specify an appropriate kernel function over the domain of interest. We provide such a choice for our application and give theoretical justification in Section 2.3.2.

2.3.2 Spherical Convolutional Wasserstein Distance

For our application to climate model validation, we specify the spatial domain $\mathcal{S}$ to be the unit sphere $\mathbb{S}^2$, the space over which latitude-longitude coordinates are assigned to locations on the Earth. To preserve local spatial information, we specify the slicing function to be a radial kernel function. We introduce the spherical convolutional WD (SCWD) as a specific case of the functional sliced WD:

Definition 2.3 (Spherical Convolutional WD)**.** Let $P, Q \in \mathcal{P}(L^2(\mathbb{S}^2))$ and $r \geq 1$. We define the (r-th order) SCWD between P and Q as follows

$$SCW_r(P,Q) = \left(\int_{\mathbb{S}^2} W_r(\omega_s \# P, \omega_s \# Q)^r ds \right)^{1/r}, \qquad \omega_s(f) = \int_{\mathbb{S}^2} f(u)\phi(s,u)du,$$

where ω_s is a convolution slicer that satisfies Definition 2.1 with associated radial kernel $\phi(s,u)$.

Because ϕ is a radial kernel function, ω_s aggregates local information and the resulting pushforward measures $\omega_s \# P$ and $\omega_s \# Q$ represent the local distribution around each location s. SCWD is therefore calculated as the global mean of the WD between local distributions at each location. The local WD values can be recorded and later visualized as a map to pinpoint regions with higher or lower similarity. Figure 2.1 demonstrates the process for calculating SCWD between two distributions of surface temperature fields, and details on the implementation are provided in Section 2.7.2.

Kernel In our analysis, we specify the kernel function to be $\phi(|s-u|; l)$, where $|s-u|$ is the chordal distance between s and u and ϕ is the Wendland kernel function used in Nychka et al. [2015]:

$$\phi(d;l) = \begin{cases} (1-d)^6(35d^2+18d+3)/3 & d \leq l, \\ 0 & d > l, \end{cases} \tag{2.2}$$

with range parameter $l > 0$ determining the radius over which the kernel is nonzero. The Wendland kernel meets the continuity assumption in Definition 2.1 and is also compact, which enables efficient sparse computations for our analysis.

Positive definite kernels, such as the Wendland kernel for l less than the diameter [Hubbert and Jäger, 2023], allow us to retain full spatial information via the spectral density. This is because the convolution theorem on $\mathbb{S}^2$ [Driscoll and Healy, 1994] gives an injective correspondence between the spectral density

of a function $f \in L^2(\mathbb{S}^2)$ and the spectral density of the convolution $(f * k)(s) = \int_{\mathbb{S}^2} f(u)k(s,u)du$ when k is positive definite. Note that as $l \to \infty$, the Wendland kernel converges to the flat kernel $\phi(d; \infty) = 1$, resulting in a SCWD where every slice is the global mean. In this case, the SCWD will be equal to the global mean-based WD from Vissio et al. [2020], leading to a complete loss of spatial variability information. In our analysis, we ensure a positive definite kernel by specifying l to be less than the diameter of the Earth (about $12,750$ km). We study the sensitivity of our results to this parameter in Section 2.4.4.

Spatial Analysis Projection selection approaches such as the Max-Sliced WD [Deshpande et al., 2019] and Energy-Based Sliced WD [Nguyen and Ho, 2024] have been introduced as alternatives to the sliced WD that give higher slicing weight to the directions of greatest variability between high-dimensional distributions. Each of our slices corresponds to a local mean around a specific location, so applying projection selection methods to SCWD would be equivalent to identifying the geographic regions in which two distributions of spatial fields have the greatest differences. This is of keen interest when validating climate models, so we examine spatial maps of our slices in Section 2.4.2 to identify these regions. However, for a comprehensive evaluation of climate fields, we favor the geographically-balanced SCWD in Definition 2.3 when evaluating similarity to reference datasets.

2.4 Climate Model Validation

We consider climate model outputs from the Coupled Model Intercomparison Project (CMIP) historical experiment phases 5 and 6. We focus on daily average near-surface (2m) temperature in degrees Celsius and daily total precipitation in mm. The CMIP6 historical simulations are organized by ensembles, each of which is distinguished with an `ripf` identifier (`rip` for CMIP5), representing realization, initialization, physics, and forcings of the model, respectively [Eyring et al., 2016a]. We obtain 46 CMIP6 model outputs with the `r1i1p1f1` ID and 33 CMIP5 model outputs with the `r1i1p1` ID. Output was obtained either at the daily frequency or aggregated from 3-hourly data. At the time of writing, two of the 79 total models did not have output available at a suitable frequency for surface temperature. All 79 models had output available for total precipitation.

To serve as references for climate model evaluation, we collect the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis 5th Generation (ERA5) [Hersbach et al., 2020] as well as the Reanalysis-2 data from the National Centers for Environmental Protection (NCEP) [Kanamitsu et al., 2002] reanalysis datasets. Both datasets were obtained for surface temperature and total precipitation at a daily frequency. Due to known issues with reanalysis data for precipitation [Tapiador et al., 2017], we obtain observations from the National Centers for Environmental Information (NCEI) Global Precipitation

Climatology Project (GPCP) Daily Precipitation Analysis Climate Data Record [Huffman et al., 2001, Adler et al., 2020] as an additional reference for precipitation.

The historical time periods for each climate variable were chosen to maximize the available model outputs and reference datasets. For surface temperature, a common time period of January 1, 1979 to November 30th, 2005 was collected for each CMIP output and reanalysis dataset. For total precipitation, we restrict the time period to October 1, 1996 to November 30th, 2005 to accommodate the first available day of observations in the GPCP dataset. Each data product represents the climate variables on a different latitude-longitude grid, which varies in size and structure. See Section 2.7.3 for full details on the spatial resolution and availability of temperature/precipitation for each dataset.

2.4.1 Evaluation of CMIP6 Models

To evaluate the skill of CMIP6 models in characterizing historical climate distributions, we compute SCWD between each model output and the reference datasets. Models which excel at replicating the local climate distribution in many different areas of the Earth will have a low SCWD. Conversely, models which fail to capture features of the local climate distributions, such as the mean, variance, or extremes, will have a higher SCWD. We do not expect perfect agreement between models and historical data, so, similar to Vissio et al. [2020], we additionally calculate SCWD between the reference datasets as a baseline for comparison. All SCWD calculations in this section use the Wendland kernel with range parameter of 1,000km for slicing.

We select the ERA5 Reanalysis as our reference dataset for (2m) surface temperature due to its high spatial resolution. We calculate distances from each surface temperature model output in our CMIP5 and CMIP6 ensembles to ERA5. In addition, we calculate SCWD between ERA5 and the NCEP Reanalysis to compare the variability between reanalysis datasets to the variability between models and reanalysis. For total precipitation, we select the GPCP observational data as our reference. We calculate distances from each total precipitation model output in our CMIP5 and CMIP6 ensembles to GPCP. We include SCWD calculations from GPCP to both ERA5 and NCEP for comparison, with the secondary goal of assessing the accuracy of each reanalysis in faithfully filling gaps in observed precipitation measurements. Full details on the SCWD rankings can be seen in Tables 2.2, 2.3, 2.4, and 2.5 in Section 2.7.4. Here, we focus on the results for CMIP6, and Figure 2.2 provides the SCWD from each CMIP6 model output to the ERA5 surface temperature field and GPCP total precipitation field.

Surface Temperature For surface temperature, NCEP has a lower SCWD to ERA5 than all CMIP models. This is not a surprise: both ERA5 and NCEP are based on observations, so we expect their temperature distributions to be similar in most regions. Among the model outputs, many models have a SCWD to

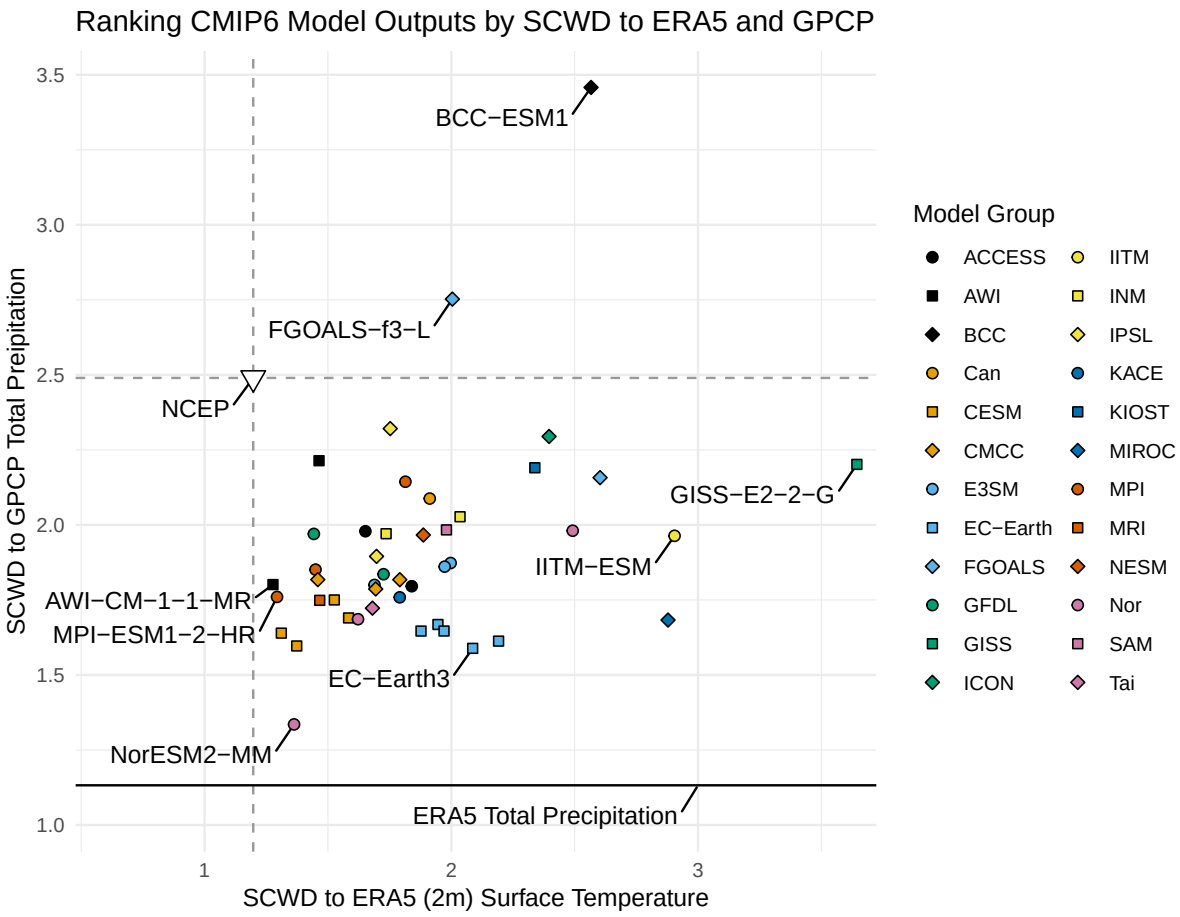


Figure 2.2: Ranking CMIP6 model outputs using SCWD. Each model output is represented by a point on the scatter plot and models from the same group share the color and shape. The x-axis and y-axis values represent the SCWD from each model output to the ERA5 surface temperature and GPCP total precipitation fields, respectively. The NCEP reanalysis is included as a white triangle with dashed lines representing the SCWD to ERA5 and GPCP. The SCWD from the ERA5 total precipitation field to GPCP is represented as a solid line.

ERA5 similar to that of NCEP. In particular, AWI-CM-1-1-MR from the Alfred Wegener Institute and MPI-ESM1-2-HR from the Max Planck Institute have the lowest SCWD for surface temperature, with a few models close behind.

Total Precipitation Compared to surface temperature, NCEP no longer has a lower SCWD to GPCP than the CMIP6 models. Instead, the ERA5 total precipitation field, which has the lowest SCWD to GPCP, serves as a better baseline for comparison. Deficiencies of precipitation from reanalysis have been reported in previous studies [Janowiak et al., 1998]. In brevity, precipitation is sensitive to model physics and is not strongly constrained by observations via data assimilation. The low SCWD of ERA5 can likely be attributed to the high model resolution and more advanced model physics and data assimilation system compared to NCEP. Among the CMIP6 models, the Norwegian Climate Centre NorESM2-MM model output stands out

with the lowest SCWD to GPCP by a relatively wide margin. The EC-Earth3 and CESM model outputs also form clusters with low SCWD values for precipitation.

No single model has the lowest SCWD value for both surface temperature and total precipitation, but NorESM2-MM seems to have the best balance of low distances for each variable and is not far off from the intersection of the lines for our reference datasets. Similarly, no model has the highest SCWD value for both climate variables. The GISS-E2-2-G model from NASA's Goddard Institute for Space Studies is a distinct outlier with a high surface temperature SCWD to ERA5, and the Beijing Climate Center BCC-ESM1 model is an outlier in terms of high SCWD to GPCP. We investigate these high SCWD values in Section 2.4.2.

2.4.2 Spatial Comparisons

Because SCWD is calculated as a global mean of local WD values, we can investigate the geographic sources of these outlying high SCWD values. Figure 2.3 provides a spatial breakdown of the local WD values obtained when calculating SCWD for surface temperature between ERA5 and AWI-CM-1-1-MR as well as ERA5 and GISS-E2-2-G, the models with the lowest and highest SCWD to ERA5, respectively. Overall, both maps seem smooth or continuous in space, with little variation between neighboring locations in most cases. For the map between AWI-CM-1-1-MR and ERA5, the local WD values are relatively low everywhere, with regions of slightly higher values near the poles and mountains. Compared to AWI-CM-1-1-MR, the GISS-E2-2-G model has similarly low WD values in the tropics. However, closer to the poles, the local WD values begin to increase. In particular, the Arctic region has extremely high WD values relative to the rest of the Earth. This is indicative of the previously documented winter cool bias in the Artic region for GISS-E2-2-G [Kelley et al., 2020].

Similar maps are provided for the NorESM2-MM and BCC-ESM1 outputs compared to GPCP total precipitation. Overall, both maps are less smooth than the surface temperature maps, potentially due to the more localized nature of precipitation. Looking at both models, the higher values of SCWD near the equator in the Pacific and Atlantic oceans may be related to the double-Intertropical Convergence Zone problem common in CMIP models, in which excessive precipitation is produced in the southern tropics [Mechoso et al., 1995]. This trend is much more pronounced for BCC-ESM1 than for NorESM2-MM. Additionally, BCC-ESM1 has a region of particularly high WD values around eastern Indonesia, where wind-terrain interaction plays an important role in the regional distribution of precipitation. This region has been previously highlighted in Zhang et al. [2021] as an area where BCC-ESM1 heavily overestimates annual mean precipitation.

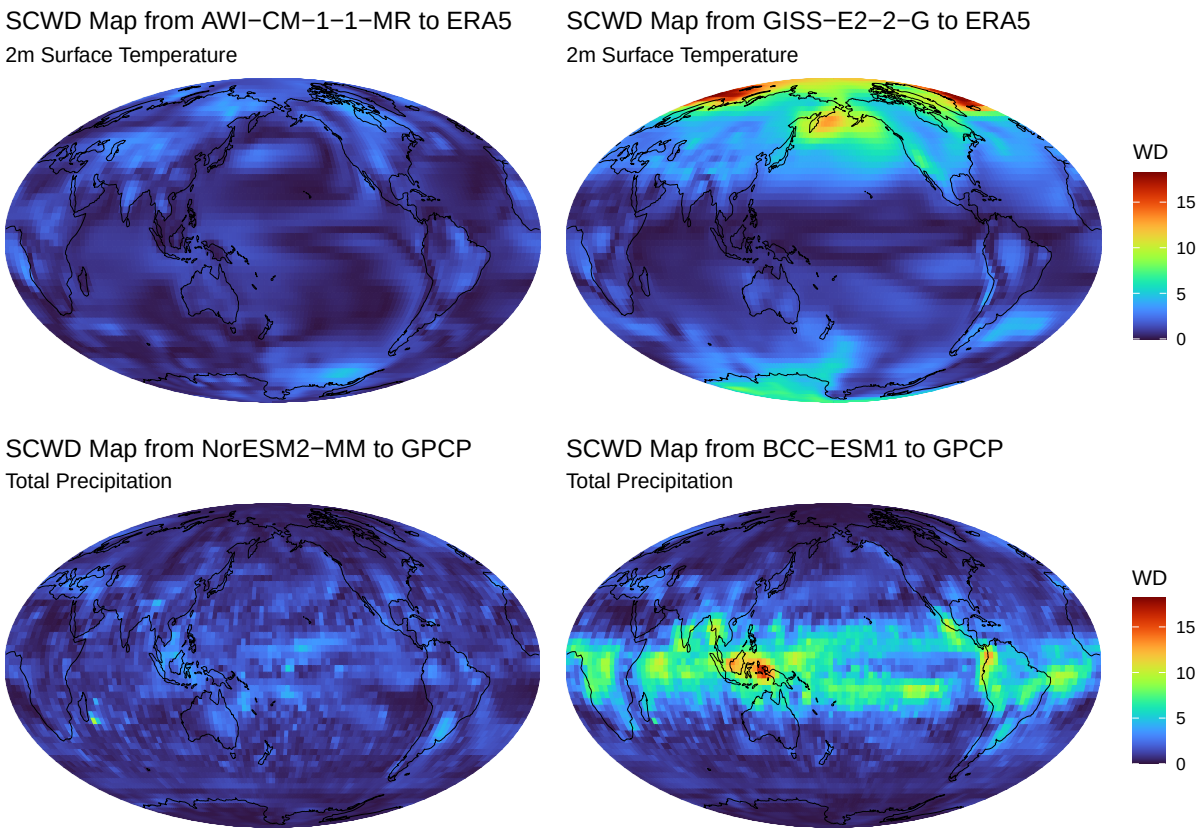


Figure 2.3: Top: Map of local Wasserstein distances from ERA5 to two CMIP6 2m surface temperature outputs: AWI-CM-1-1-MR and GISS-E2-2-G. Bottom: Map of local Wasserstein distances from GPCP to two CMIP6 total precipitation outputs: NorESM2-MM and BCC-ESM1. Color fill at each location is determined by the WD between the local distributions obtained from the convolution slicer in Definition 2.3. The color scale is shared for all maps and continental boundaries are included in black to aid spatial comparisons.

2.4.3 Comparing CMIP5 and CMIP6

To assess the progression from CMIP5 to CMIP6, Figure 2.4 provides boxplots of the SCWD from each CMIP model to the reference datasets. The left panel provides SCWD calculations from the surface temperature models and NCEP to the ERA5 Reanalysis. The median SCWD value for CMIP6 models to ERA5 is lower than that of CMIP5. However, the two boxplots share a similar range, so the difference between the CMIP5 and CMIP6 ensembles is subtle. Overall, we see a promising, albeit limited, decrease in SCWD for typical CMIP6 models compared to CMIP5. This indicates improved performance of CMIP6 when it comes to reconstructing realistic temperature distributions at the local level. The right panel provides SCWD calculations from each precipitation model and ERA5/NCEP to the GPCP dataset. The median SCWD value for CMIP6 is again lower than that of CMIP5. Compared to surface temperature, the difference between CMIP5 and CMIP6 is more distinct. This indicates that relative to surface temperature, precipitation modeling has seen greater gains with the transition from CMIP5 to CMIP6. The improvement of CMIP6

models compared to CMIP5 in precipitation representation has been reported in previous studies using different evaluation methods (e.g., Chen et al. [2021]). It can be probably attributed to the more advanced model physics and overall higher model resolution in CMIP6.

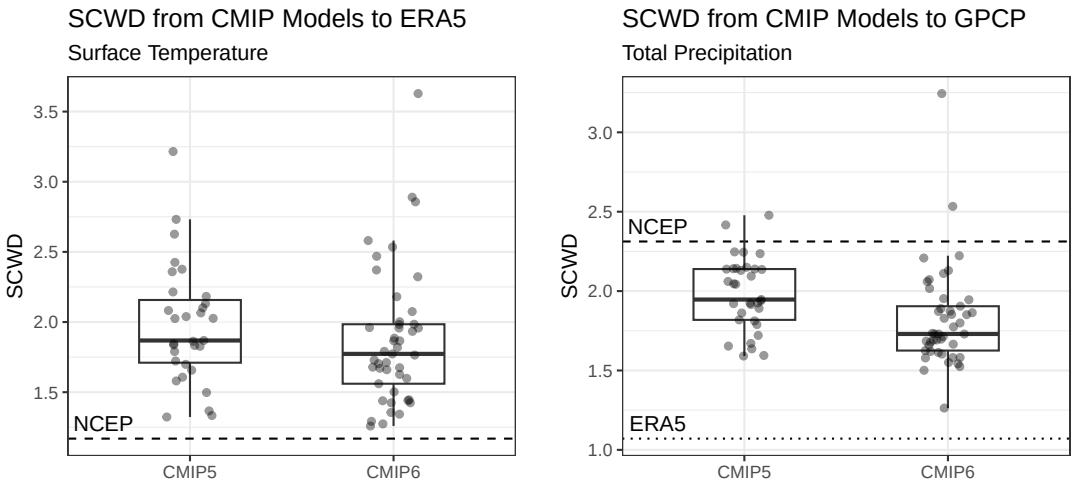


Figure 2.4: Left: Boxplots of SCWD from the CMIP5 and CMIP6 model outputs to the ERA5 Reanalysis for 2m surface temperature. Right: Boxplots of SCWD from the CMIP5 and CMIP6 model outputs to the GPCP observational dataset for total precipitation. Each plot contains points representing the SCWD from each CMIP model output to the reference dataset (ERA5 or GPCP), with CMIP5 and CMIP6 separated into two boxplots for comparison. Dotted and dashed lines are included to represent the SCWD from NCEP to ERA5/GPCP and ERA5 to GPCP, respectively.

2.4.4 Metric Comparison

We compare our method to baseline methods and evaluate the sensitivity of our results to the choice of range parameter. We generate seven additional rankings for the CMIP models: SCWD with a range of 500km and 2,500km respectively, the global mean-based WD (GMWD) from Vissio et al. [2020], RMSE and MAE computed on long-term mean climatologies, and WD and Sliced WD on data regridded to an icosahedral grid to preserve area weighting. Technical details and rankings together with our original 1,000km SCWD results can be seen in Tables 2.2 and 2.3 in Section 2.7.4.

All SCWD rankings are nearly identical regardless of the range parameter, except that several precipitation models, such as GFDL-CM4 and E3SM-2-0, rank better with a larger range. From here forward, we focus only on the chosen 1,000km range parameter for SCWD. For most baseline methods (RMSE, MAE, WD, and Sliced WD), their rankings are moderately similar to those of SCWD, indicating general agreements, though each metric still shows unique perspectives. However, there is a large discrepancy between the rankings of GMWD and all other rankings. For surface temperature, SCWD, RMSE, and WD all show that the distance from NCEP to ERA5 is lower than that of all CMIP6 models, and MAE and Sliced WD also rank NCEP

favorably among the models. However, GMWD ranks NCEP only among the median CMIP6 models. For precipitation, we see a similar result for ERA5 compared to GPCP: ERA5 has the best ranking when using SCWD, RMSE, MAE, and WD, but is third in the rankings for Sliced WD and again is around the median for GMWD.

To better understand the differences in rankings between SCWD and baseline model evaluation methods such as GMWD and RMSE/MAE, Section 2.7.5 provides an experiment and data example. The experiment shows a case where SCWD detects changes in both the climatological mean and the variance of the anomalies, while the other methods detect only one type of change. In particular, we show that the RMSE and MAE criteria are unable to detect isolated changes in the variance of the anomalies. The data example focuses on the SAM0-UNICON surface temperature model. Despite the overall similarity in rankings for CMIP6 surface temperature, this model ranks significantly higher by RMSE/MAE than by SCWD. Further investigation shows that SAM0-UNICON exhibits significant differences in both the climatological mean and the variance of the anomalies. Because RMSE/MAE are unable to detect shifts in the variance of the anomalies they artificially inflate the ranking of this model compared to SCWD.

2.5 Discussion

Climate model validation is critical for ensuring that climate models faithfully represent the Earth system. To this end, we developed a new similarity measure, called spherical convolutional Wasserstein distance (SCWD), which quantifies model performance in a way that properly accounts for spatial variability. SCWD builds on previous sliced WD methods to compare distributions of infinite-dimensional functional data, specifically surfaces on the sphere $\mathbb{S}^2$. Overall, our results indicate that incorporating local perspectives, rather than just the global mean, is essential when evaluating the similarities between climate models and observational data. In both theory and practice, a model that represents the global mean well may have large compensating errors at the regional scale. SCWD better represents the regional performance of climate models than the previously proposed WD-based evaluation criteria in [Vissio et al., 2020]. Furthermore, SCWD is not limited to only comparing the long-term mean climate state as with RMSE and MAE, and our convolution slicing provides better spatial insights (in the form of maps) than WD and Sliced WD.

For surface temperature, we found evidence to suggest that SCWD is more accurate than the previous WD-based approach for evaluating climate models, such as the two reanalysis datasets being more similar compared to climate models and reanalysis. However, for precipitation, similar findings were made for only one of the reanalysis datasets. Given the limitations with reanalysis for precipitation discussed in Section 2.4.1, additional experiments from different perspectives may be needed to fully assess the accuracy of our

method. Because we have only shown that SCWD is a pseudometric, rather than a metric, SCWD may not be able to fully distinguish between climate fields which have differences only in the joint spatial distribution, rather than at the local level. We acknowledge that determining the sample complexity of SCWD is an area that requires further investigation.

We hope our method will be tested under many different scenarios in climate science. For example, calculating SCWD on monthly mean or even annual data could help to detect biases in terms of longer range temporal variability. Another idea is to remove long-term climatological means from the data before computing SCWD. This would provide additional insight into which features are being captured beyond the mean climate state. Outside of climate science, SCWD could be directly applied to compare distributions of 360^o images. In this case, a straightforward extension would allow for multiple convolution layers similar to the rectangular case in Nguyen and Ho [2022]. SCWD could also be adapted to compare distributions of spatiotemporal fields, rather than spatial fields as in this article, by including both time and space in the functional data domain. This would require specifying a space-time kernel function [Porcu et al., 2016]. The more general functional sliced WD framework has a variety of potential use cases for comparing distributions of functions on a broad class of manifolds. Potential application areas where the data lie on non-trivial manifolds include facial recognition [Li et al., 2014], astronomy [Szapudi, 2008], and ecology [Sutherland et al., 2015].

2.6 Acknowledgements

We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. NCEP/DOE Reanalysis II data provided by the NOAA PSL, Boulder, Colorado, USA, from their website at `https://psl.noaa.gov`. This work is partially supported by NSF-DMS-1830312, NSF-DGE-1922758, and NSF-DMS-2124576 from the National Science Foundation.

2.7 Supplementary Material

2.7.1 Proof of Theorem 2.1

Proof. Let $r \geq 1$, let $\mathcal{S}$ be a compact subset of $\mathbb{R}^n$, and let $P, Q, U \in \mathcal{P}(L^2(\mathcal{S}))$. We show the identity property holds:

$$\begin{aligned} FSW_r(P,P) &= \left(\int_{\mathcal{S}} W_r(c_s\#P, c_s\#P)^r ds \right)^{1/r} \\ &= \left(\int_{\mathcal{S}} 0^r ds \right)^{1/r} \\ &= 0. \end{aligned}$$

The second line holds by the identity property of the ordinary WD. Next, the symmetry property:

$$\begin{aligned} FSW_r(P,Q) &= \left(\int_{\mathcal{S}} W_r(c_s\#P, c_s\#Q)^r ds \right)^{1/r} \\ &= \left(\int_{\mathcal{S}} W_r(c_s\#Q, c_s\#P)^r ds \right)^{1/r} \\ &= FSW_r(Q,P). \end{aligned}$$

The second line holds by the symmetry property of the ordinary WD. Next, the triangle inequality:

$$\begin{aligned} FSW_r(P,U) &= \left(\int_{\mathcal{S}} W_r(c_s\#P, c_s\#U)^r ds \right)^{1/r} \\ &\leq \left(\int_{\mathcal{S}} \left[W_r(c_s\#P, c_s\#Q) + W_r(c_s\#Q, c_s\#U) \right]^r ds \right)^{1/r} \\ &\leq \left(\int_{\mathcal{S}} W_r(c_s\#P, c_s\#Q)^r ds \right)^{1/r} + \left(\int_{\mathcal{S}} W_r(c_s\#Q, c_s\#U)^r ds \right)^{1/r} \\ &= FSW_r(P,Q) + FSW_r(Q,U). \end{aligned}$$

The second line holds by the triangle inequality property of W_r and the third line holds by the Minkowski inequality. Lastly, we check the r-convexity property. Let $\lambda \in [0,1]$. If we take P as the reference distribution,

this property tells us what happens to the functional sliced WD as we interpolate between Q and U:

$$\begin{aligned}
FSW_r(P, \lambda Q + (1-\lambda)U) &= \left(\int_{\mathcal{S}} W_r(c_s\#P, c_s\#[\lambda Q + (1-\lambda)U])^r ds \right)^{1/r} \\
&= \left(\int_{\mathcal{S}} W_r(c_s\#P, \lambda c_s\#Q + (1-\lambda)c_s\#U)^r ds \right)^{1/r} \\
&\leq \left(\int_{\mathcal{S}} \lambda W_r(c_s\#P, c_s\#Q)^r + (1-\lambda) W_r(c_s\#P, c_s\#U)^r ds \right)^{1/r} \\
&\leq \left(\lambda \int_{\mathcal{S}} W_r(c_s\#P, \lambda c_s\#Q)^r ds \right)^{1/r} + \\
&\quad \left((1-\lambda) \int_{\mathcal{S}} W_r(c_s\#P, c_s\#U)^r ds \right)^{1/r} \\
&= \lambda^{1/r} FSW_r(P, Q) + (1-\lambda)^{1/r} FSW_r(P, U).
\end{aligned}$$

The second line follows from properties of linear operators (in this case c_s). The third line follows by the r-convexity of the ordinary WD, and the fourth line follows from the Minkowski inequality. We have shown all three pseuduometric properties and the $r-$convexity property, so our proof is complete. □

2.7.2 SCWD Implementation

Algorithm 1 describes the process by which we calculated the SCWD values shown in Section 2.4. All computations were performed using R version 4.3.2 on an Ubuntu operating system with an Intel i5-9600k processor (6 cores), 32GB RAM, and 3TB hard disk space. Total computation time for all experiments was under 72 hours. The calculation of each distance measure was fast, but loading the required datasets into memory for processing took the majority of the compute time.

The climate model outputs and reanalysis datasets (ERA, NCEP) used in our analysis have no missing data. However, the GPCP observational dataset used as the reference for total precipitation did have missing data at some sites for a few days in the historical period. To handle missing data in the GPCP dataset, we modified the slicing process in Step 2. If locations corresponding to greater than 50% of the convolution weight were missing when calculating a slice value, an `NA` value was recorded. The local WD calculations in Step 3 were computed ignoring these `NA` values. In total, 17,943 slices were missing sufficient data when using the 50% threshold. However, a total of 24,105,600 slices were considered for the GPCP dataset (3,348 days in the analysis period with 7,200 slices per day in the strided convolution), so the missing slices constituted under 0.1% of the final sliced GPCP data used in the analysis.

Algorithm 1 Spherical Convolutional Wasserstein Distance Approximation

DATA

Reference dataset $X_0(t), t \in \mathcal{T}_0$: sample of spatial fields

Model outputs $X_1(t), t \in \mathcal{T}_1, ..., X_n(t), t \in \mathcal{T}_n$: n samples of spatial fields to be compared to X_0

PARAMETERS AND APPROXIMATION GRIDS

Wasserstein order parameter r: 2

Range parameter l: 1,000 km kernel radius

Approximation quantiles Q: 200 evenly spaced quantiles ranging from 0 to 1 with a step size of 0.005

Grid G_1: 60×120 regular latitude-longitude grid of center points for strided convolution

Grid G_2: 361×720 regular latitude-longitude grid to provide discrete approximation of spherical domain

STEP 1. PRECOMPUTE SLICING WEIGHTS

for each location $s \in G_1$ **do**

Calculate vector of chordal distances from s to all locations in G_2

Calculate Wendland function (2.2) with range l for all locations in G_2 using distance vector

Apply area weighting to the Wendland kernel values using the area of each grid cell in G_2

Normalize area-weighted kernel and store the results as a sparse vector $W(s)$

end for

STEP 2. COMPUTE SLICED QUANTILES

for $i \in 0, 1, ..., n$ **do**

Re-grid X_i to G_2 without smoothing (one nearest neighbor upsampling)

for each location $s \in G_1$ **do**

Slice X_i into one dimension using the dot product $X_i^*(s,t) = \langle W(s), X_i(t) \rangle$

Calculate the sliced quantile function of $X_i^*(s,t)$, denoted as $F_i^{-1}(s,q), q \in Q$

end for

end for

STEP 3. CALCULATE APPROXIMATE SCWD

for $i \in 1, ..., n$ **do**

for each location $s \in G_1$ **do**

Calculate the local WD between X_0 and X_i centered around s as $d_i(s)^r = \sum_{q \in Q} |F_0^{-1}(q) - F_i^{-1}(q)|^r$

end for

Calculate the approximate SCWD between X_0 and X_i as $SCWD(X_0, X_i) \approx \left(\sum_{s \in G_1} d_i(s)^r \right)^{1/r}$

end for

2.7.3 Data Details and Accessibility

Obs./Reanalysis Data	Longs	Lats	TAS	PR
NCEP Reanalysis	144	73	Yes	Yes
ERA5 Reanalysis	1440	721	Yes	Yes
GPCP Observations	360	180	No	Yes

CMIP5 Models	Longs	Lats	TAS	PR
ACCESS1-0	192	145	Yes	Yes
ACCESS1-3	192	145	Yes	Yes
CanCM4	128	64	No	Yes
CMCC-CESM	96	48	Yes	Yes
CMCC-CM	480	240	Yes	Yes
CMCC-CMS	192	96	Yes	Yes
CNRM-CM5	256	128	Yes	Yes
CSIRO-Mk3-6-0	192	96	Yes	Yes
CanESM2	128	64	Yes	Yes
EC-EARTH	320	160	Yes	Yes
FGOALS-g2	128	60	Yes	Yes
FGOALS-s2	128	108	Yes	Yes
GFDL-CM3	144	90	Yes	Yes
GFDL-ESM2G	144	90	Yes	Yes
GFDL-ESM2M	144	90	Yes	Yes
HadCM3	96	73	Yes	Yes
HadGEM2-AO	192	145	Yes	Yes
HadGEM2-CC	192	145	Yes	Yes
HadGEM2-ES	192	145	Yes	Yes
INMCM4	180	120	Yes	Yes
IPSL-CM5A-LR	96	96	Yes	Yes
IPSL-CM5A-MR	144	143	Yes	Yes
IPSL-CM5B-LR	96	96	Yes	Yes
MIROC-ESM	128	64	Yes	Yes
MIROC-ESM-CHEM	128	64	Yes	Yes
MIROC4h	640	320	Yes	Yes
MIROC5	256	128	Yes	Yes
MPI-ESM-LR	192	96	Yes	Yes
MPI-ESM-MR	192	96	Yes	Yes
MPI-ESM-P	192	96	Yes	Yes
MRI-CGCM3	320	160	Yes	Yes
MRI-ESM1	320	160	Yes	Yes
NorESM1-M	144	96	Yes	Yes

CMIP6 Models	Longs	Lats	TAS	PR
ACCESS-CM2	192	144	Yes	Yes
ACCESS-ESM1-5	192	145	Yes	Yes
AWI-CM-1-1-MR	384	192	Yes	Yes
AWI-ESM-1-1-LR	192	96	Yes	Yes
BCC-ESM1	128	64	Yes	Yes
CESM2	288	192	Yes	Yes
CESM2-FV2	144	96	Yes	Yes
CESM2-WACCM	288	192	Yes	Yes
CESM2-WACCM-FV2	144	96	Yes	Yes
CMCC-CM2-HR4	288	192	Yes	Yes
CMCC-CM2-SR5	288	192	Yes	Yes
CMCC-ESM2	288	192	Yes	Yes
CanESM5	128	64	Yes	Yes
E3SM-1-0	360	180	Yes	Yes
E3SM-2-0	360	180	Yes	Yes
E3SM-2-0-NARRM	360	180	Yes	Yes
EC-Earth3	512	256	Yes	Yes
EC-Earth3-AerChem	512	256	Yes	Yes
EC-Earth3-CC	512	256	Yes	Yes
EC-Earth3-Veg	512	256	Yes	Yes
EC-Earth3-Veg-LR	320	160	Yes	Yes
FGOALS-f3-L	288	180	Yes	Yes
FGOALS-g3	180	80	Yes	Yes
GFDL-CM4	288	180	Yes	Yes
GFDL-ESM4	288	180	Yes	Yes
GISS-E2-2-G	144	90	Yes	Yes
ICON-ESM-LR*	N/A	N/A	Yes	Yes
IITM-ESM	192	94	Yes	Yes
INM-CM4-8	180	120	Yes	Yes
INM-CM5-0	180	120	Yes	Yes
IPSL-CM5A2-INCA	96	96	Yes	Yes
IPSL-CM6A-LR	144	143	Yes	Yes
IPSL-CM6A-LR-INCA	144	143	No	Yes
KACE-1-0-G	192	144	Yes	Yes
KIOST-ESM	192	96	Yes	Yes
MIROC6	256	128	Yes	Yes
MPI-ESM-1-2-HAM	192	96	Yes	Yes
MPI-ESM1-2-HR	384	192	Yes	Yes
MPI-ESM1-2-LR	192	96	Yes	Yes
MRI-ESM2-0	320	160	Yes	Yes
NESM3	192	96	Yes	Yes
NorCPM1	144	96	Yes	Yes
NorESM2-LM	144	96	Yes	Yes
NorESM2-MM	288	192	Yes	Yes
SAM0-UNICON	288	192	Yes	Yes
TaiESM1	288	192	Yes	Yes

Table 2.1: Details for each observed/reanalysis data product and CMIP model output. The columns give the model name, longitude and latitude resolution, and availability of (2m) surface temperature (TAS) and total precipitation (PR) for each dataset. All datasets were obtained on a rectangular grid except ICON-ESM-LR, which was obtained on an icosahedral grid with 10,242 total cells.

The CMIP5 and CMIP6 outputs are available at: https://esgf-node.llnl.gov/projects/esgf-llnl/

The ERA5 hourly data on single levels from 1940 to present is available at: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview

The NCEP/DOE Reanalysis II is available at: https://psl.noaa.gov/data/gridded/data.ncep.

reanalysis2.html

The GPCP 1 Degree Daily Precipitation Estimate is available at: https://www.ncei.noaa.gov/products/climate-data-records/precipitation-gpcp-daily

2.7.4 Full SCWD Rankings and Metric Comparison

In the following pages, we provide tables detailing the climate model rankings generated by SCWD and baseline methods from both the climate science and ML literature. First, we provide details on the implementation of the baseline metrics. See Section 2.4.4 for a discussion of these results.

Baseline Climate Model Evaluation Metrics Past metrics for climate model validation include the root mean square error (RMSE) and mean absolute error (MAE) applied to the climatologies, or long term means of the data [Gleckler et al., 2016]. To implement these approaches, we first require regridding the data to a common dimension. So, we regrid the data to a common resolution of 120×60 using the `remapnn` function from the Climate Data Operator (CDO) command line tools. This resolution was chosen to match the grid of locations over which the slices were computed for SCWD. Afterwards, we compute long term means at each location, ignoring leap days, to produce 365 values representing the climatologies for each day of the year. RMSE and MAE are computed between the climatologies for the reference dataset and each climate model at each location. The final rankings are determined by the area-weighted global mean over RMSE/MAE values at each location. Models marked as "NA" operate on a 360-day calendar, making it challenging to estimate a 365-day climatology to match the reference datasets.

Baseline Wasserstein Distance Metrics We compare our approach against existing Wasserstein-based metrics. Although these metrics have never been used for model validation, they can still provide a sense of what our method offers over the naive application of existing tools. We include ordinary Wasserstein distance (WD), Sliced Wasserstein Distance (SWD), and attempted to include Convolution SWD (CSWD). For these methods to work, we first have to regrid each model to a common set of grid points. However, unlike the baseline RMSE and MAE-based approaches described in the previous paragraph, none of the proposed Wasserstein methods account for area weighting of geographic data. So, we again regrid the climate models and reference datasets to a common resolution but now using an Icosahedral grid of 642 locations, which provides near-uniform spacing on the sphere to alleviate concerns about area weighting. The reduced dimensionality is not ideal for evaluating climate models, but serves to make the computation of the multivariate WD and sliced WD more feasible given the costs already incurred from regridding. Because we are using an Icosahedral grid to account for area weighting, we are unable to compute convolution sliced WD, which is designed for square grids.

CMIP6 Surface Temperature Rankings

	SCWD (500km)	SCWD (1000km)	SCWD (2500km)	GMWD	RMSE (climatology)	MAE (climatology)	WD (icosahedral)	SWD (icosahedral)
NCEP Reanalysis	1.412	1.197	0.923	0.308	1.456	1.341	63.307	1.455
AWI−CM−1−1−MR	1.382	1.277	1.044	0.308	1.518	1.309	86.069	1.214
MPI−ESM1−2−HR	1.404	1.294	1.039	0.389	1.582	1.368	85.374	1.322
CESM2−WACCM	1.419	1.311	1.074	0.277	1.495	1.288	85.684	1.276
NorESM2−MM	1.476	1.363	1.121	0.095	1.468	1.262	86.775	1.399
CESM2	1.476	1.373	1.149	0.449	1.515	1.307	85.850	1.321
GFDL−ESM4	1.544	1.443	1.196	0.377	1.496	1.263	87.304	1.289
MPI−ESM1−2−LR	1.600	1.450	1.117	0.070	1.706	1.483	89.824	1.539
CMCC−CM2−HR4	1.561	1.459	1.228	0.722	1.714	1.506	89.084	1.409
AWI−ESM−1−1−LR	1.614	1.464	1.150	0.476	1.792	1.573	91.583	1.544
MRI−ESM2−0	1.581	1.467	1.198	0.161	1.613	1.374	92.149	1.510
CESM2−WACCM−FV2	1.679	1.526	1.246	0.452	1.713	1.489	91.093	1.641
CESM2−FV2	1.734	1.584	1.296	0.331	1.729	1.501	92.046	1.658
NorESM2−LM	1.774	1.622	1.341	0.645	1.828	1.607	93.452	1.819
ACCESS−ESM1−5	1.816	1.652	1.365	0.767	1.922	1.696	93.122	1.621
TaiESM1	1.797	1.680	1.443	0.141	1.601	1.381	91.811	1.606
E3SM−1−0	1.797	1.689	1.456	0.182	1.773	1.522	93.189	1.558
CMCC−ESM2	1.802	1.694	1.455	0.592	1.645	1.396	90.681	1.548
IPSL−CM6A−LR	1.803	1.697	1.458	0.557	1.709	1.454	94.189	1.618
GFDL−CM4	1.806	1.725	1.552	1.008	1.659	1.430	89.509	1.379
INM−CM5−0	1.937	1.736	1.364	0.435	1.977	1.724	98.534	1.832
IPSL−CM5A2−INCA	1.895	1.752	1.450	0.446	1.992	1.718	94.499	1.676
KACE−1−0−G	1.960	1.791	1.499	0.269	NA	NA	96.472	1.835
CMCC−CM2−SR5	1.896	1.791	1.559	0.702	1.721	1.463	91.996	1.677
MPI−ESM−1−2−HAM	1.949	1.814	1.521	0.174	1.897	1.668	93.734	1.661
ACCESS−CM2	1.964	1.840	1.597	0.092	1.868	1.638	94.319	1.685
EC−Earth3−Veg	1.943	1.877	1.737	0.475	1.807	1.596	93.386	1.695
NESM3	2.011	1.887	1.623	0.133	2.097	1.774	96.014	1.844
CanESM5	2.104	1.913	1.567	0.265	2.123	1.877	102.338	2.106
EC−Earth3−AerChem	2.016	1.946	1.795	0.207	1.868	1.656	94.194	1.738
EC−Earth3−Veg−LR	2.051	1.971	1.817	0.150	1.993	1.768	96.038	1.861
E3SM−2−0	2.072	1.973	1.775	0.541	1.999	1.748	96.480	1.820
SAM0−UNICON	2.091	1.980	1.747	0.744	1.718	1.486	96.725	1.725
E3SM−2−0−NARRM	2.087	1.997	1.815	0.658	1.990	1.755	96.134	1.740
FGOALS−f3−L	2.119	2.004	1.751	0.614	1.886	1.635	96.587	1.703
INM−CM4−8	2.236	2.035	1.629	0.252	2.086	1.821	101.024	1.917
EC−Earth3	2.156	2.087	1.939	0.238	1.955	1.738	97.007	1.812
EC−Earth3−CC	2.254	2.192	2.054	0.842	1.972	1.761	96.699	1.810
KIOST−ESM	2.430	2.338	2.139	0.631	2.178	1.927	102.481	2.065
ICON−ESM−LR	2.519	2.396	2.010	0.489	2.416	2.095	103.412	2.280
NorCPM1	2.638	2.492	2.195	1.218	2.301	2.025	108.400	2.329
BCC−ESM1	2.774	2.566	2.212	0.905	2.680	2.422	110.310	2.560
FGOALS−g3	2.738	2.603	2.325	0.695	2.180	1.918	108.211	2.479
MIROC6	3.006	2.878	2.611	1.531	2.360	2.103	108.782	2.423
IITM−ESM	2.996	2.905	2.709	0.692	2.496	2.258	110.685	2.525
GISS−E2−2−G	3.726	3.644	3.432	1.832	2.802	2.533	122.535	2.851

Table 2.2: CMIP6 model rankings for (2m) surface temperature based on similarity to the ERA5 Reanalysis. Distances are calculated using our proposed spherical convolutional WD (SCWD) as well as the global mean-based WD (GMWD), RMSE and MAE (both computed on the climatologies), WD and Sliced WD (both computed on data regridded to an icosahedral grid). For the SCWD calculations, three different range parameters are chosen for the Wendland kernel: 500km, 1000km (our proposed choice), and 2500km. Color fill is unique to each column in the table, and is calculated using ranks.

CMIP6 Total Precipitation Rankings

	SCWD (500km)	SCWD (1000km)	SCWD (2500km)	GMWD	RMSE (climatology)	MAE (climatology)	WD (icosahedral)	SWD (icosahedral)
ERA5	1.804	1.133	0.701	0.252	1.936	1.349	155.514	0.928
NorESM2−MM	1.905	1.335	0.762	0.167	2.579	1.845	182.057	0.766
EC−Earth3	2.028	1.589	0.968	0.236	2.608	1.906	175.419	1.136
CESM2	2.425	1.597	0.891	0.263	2.678	1.930	184.416	0.945
EC−Earth3−CC	2.043	1.613	0.998	0.275	2.660	1.945	177.141	1.159
CESM2−WACCM	2.399	1.640	0.889	0.246	2.648	1.909	182.414	0.903
EC−Earth3−Veg	2.083	1.646	1.009	0.249	2.656	1.936	176.401	1.183
EC−Earth3−Veg−LR	2.080	1.647	1.006	0.193	2.618	1.902	172.882	1.264
EC−Earth3−AerChem	2.120	1.668	1.031	0.241	2.656	1.939	176.499	1.128
MIROC6	2.592	1.683	0.946	0.486	2.840	2.048	195.362	1.294
NorESM2−LM	2.137	1.686	0.972	0.182	2.674	1.916	180.257	1.051
CESM2−FV2	2.134	1.690	1.014	0.265	2.648	1.910	176.164	1.129
TaiESM1	2.593	1.723	1.028	0.384	2.652	1.948	180.010	0.995
MRI−ESM2−0	2.453	1.749	1.044	0.304	2.767	2.014	186.046	1.148
CESM2−WACCM−FV2	2.199	1.750	1.037	0.262	2.696	1.948	176.087	1.249
KACE−1−0−G	2.789	1.759	1.044	0.340	NA	NA	203.446	1.868
MPI−ESM1−2−HR	2.288	1.760	1.040	0.215	2.765	1.994	183.813	1.140
CMCC−ESM2	2.602	1.787	1.073	0.402	2.675	1.969	179.970	1.058
ACCESS−CM2	2.816	1.796	1.062	0.447	3.059	2.179	209.684	2.051
E3SM−1−0	2.432	1.800	1.055	0.359	2.715	1.976	178.568	1.154
AWI−CM−1−1−MR	2.327	1.802	1.074	0.234	2.769	2.009	185.090	1.288
CMCC−CM2−SR5	2.578	1.818	1.115	0.412	2.710	1.986	179.817	1.104
CMCC−CM2−HR4	2.601	1.818	1.112	0.291	2.827	2.015	186.971	1.103
GFDL−CM4	3.064	1.836	0.917	0.225	2.832	1.980	199.471	1.425
MPI−ESM1−2−LR	2.353	1.851	1.093	0.167	2.726	1.973	177.289	1.380
E3SM−2−0	3.042	1.861	0.967	0.276	2.736	1.942	184.789	1.033
E3SM−2−0−NARRM	3.120	1.873	0.953	0.276	2.732	1.943	187.589	1.030
IPSL−CM6A−LR	2.845	1.895	1.143	0.345	2.948	2.098	203.981	1.700
IPSL−CM6A−LR−INCA	2.926	1.952	1.163	0.349	2.963	2.101	204.900	1.800
IITM−ESM	2.616	1.964	1.177	0.240	2.937	2.115	188.449	1.209
NESM3	2.446	1.967	1.196	0.181	2.937	2.073	184.775	1.083
GFDL−ESM4	3.194	1.970	1.008	0.271	2.932	2.040	202.174	1.601
INM−CM5−0	2.973	1.971	1.207	0.359	2.844	2.069	190.630	1.230
ACCESS−ESM1−5	2.827	1.979	1.158	0.546	2.965	2.152	197.594	1.538
NorCPM1	2.469	1.980	1.233	0.080	2.701	1.935	172.650	1.530
SAM0−UNICON	3.064	1.984	1.066	0.332	2.797	2.015	187.960	1.094
INM−CM4−8	2.658	2.027	1.333	0.368	2.790	2.061	180.967	1.307
CanESM5	3.303	2.088	1.123	0.209	2.973	2.090	198.500	1.557
MPI−ESM−1−2−HAM	2.681	2.144	1.294	0.238	2.902	2.100	181.616	1.552
FGOALS−g3	3.438	2.157	1.136	0.097	3.136	2.208	219.320	2.460
KIOST−ESM	3.008	2.190	1.239	0.091	2.757	1.983	179.854	1.222
GISS−E2−2−G	2.847	2.201	1.366	0.152	2.857	2.092	180.662	1.519
AWI−ESM−1−1−LR	2.745	2.214	1.349	0.160	2.800	2.035	176.481	1.739
ICON−ESM−LR	2.849	2.295	1.413	0.118	3.016	2.153	185.359	1.428
IPSL−CM5A2−INCA	2.996	2.321	1.380	0.143	2.796	2.039	173.032	1.766
NCEP	3.738	2.490	1.289	0.581	2.758	1.943	215.618	2.533
FGOALS−f3−L	4.614	2.752	1.264	0.191	3.155	2.199	226.689	2.839
BCC−ESM1	4.575	3.458	1.491	0.110	3.141	2.156	219.888	2.436

Table 2.3: CMIP6 model rankings for total precipitation based on similarity to the GPCP observations. Distances are calculated using our proposed spherical convolutional WD (SCWD) as well as the global mean-based WD (GMWD), RMSE and MAE (both computed on the climatologies), WD and Sliced WD (both computed on data regridded to an icosahedral grid). For the SCWD calculations, three different range parameters are chosen for the Wendland kernel: 500km, 1000km (our proposed choice), and 2500km. Color fill is unique to each column in the table, and is calculated using ranks.

CMIP5 Surface Temperature Rankings

	SCWD (500km)	SCWD (1000km)	SCWD (2500km)	GMWD	RMSE (climatology)	MAE (climatology)	WD (icosahedral)	SWD (icosahedral)
NCEP Reanalysis	1.412	1.197	0.923	0.308	1.456	1.341	63.307	1.455
MPI−ESM−P	1.491	1.347	1.039	0.157	1.671	1.440	88.946	1.505
MPI−ESM−MR	1.491	1.356	1.070	0.292	1.661	1.431	89.364	1.506
MPI−ESM−LR	1.530	1.390	1.089	0.132	1.687	1.458	89.557	1.502
MIROC4h	1.620	1.516	1.292	0.648	1.638	1.408	85.834	1.417
ACCESS1−3	1.807	1.612	1.238	0.209	1.826	1.593	93.936	1.735
NorESM1−M	1.816	1.634	1.310	0.485	1.855	1.623	94.197	1.775
ACCESS1−0	1.848	1.683	1.385	0.055	1.780	1.531	94.300	1.633
IPSL−CM5A−MR	1.833	1.719	1.448	0.307	1.869	1.637	91.688	1.643
GFDL−CM3	1.877	1.745	1.446	0.371	1.884	1.634	92.071	1.626
HadGEM2−AO	1.977	1.815	1.516	0.305	NA	NA	96.722	1.692
EC−EARTH	1.921	1.841	1.650	0.862	1.994	1.777	93.500	1.720
HadGEM2−ES	2.014	1.857	1.582	0.336	NA	NA	97.603	1.845
CMCC−CMS	1.984	1.865	1.617	0.157	1.999	1.767	97.743	1.813
MIROC−ESM−CHEM	2.118	1.870	1.528	0.322	2.162	1.902	98.743	2.045
CNRM−CM5	2.018	1.891	1.618	0.303	1.986	1.720	97.237	1.828
MIROC−ESM	2.141	1.896	1.559	0.276	2.170	1.905	98.937	2.054
GFDL−ESM2M	2.203	2.049	1.763	0.192	2.189	1.909	100.727	1.912
CanESM2	2.267	2.057	1.703	0.265	2.188	1.914	104.647	2.213
IPSL−CM5A−LR	2.181	2.059	1.805	1.111	2.180	1.936	97.748	1.977
MRI−ESM1	2.187	2.084	1.848	0.108	2.088	1.808	98.953	1.946
CMCC−CM	2.197	2.100	1.861	0.273	1.981	1.746	99.122	1.942
MRI−CGCM3	2.224	2.120	1.881	0.167	2.090	1.806	99.788	1.965
GFDL−ESM2G	2.309	2.155	1.868	0.348	2.273	1.987	101.759	1.975
HadGEM2−CC	2.350	2.205	1.950	0.628	NA	NA	102.304	2.205
MIROC5	2.357	2.235	1.972	0.855	2.013	1.787	96.780	1.965
CMCC−CESM	2.614	2.389	2.088	0.147	2.549	2.280	109.000	2.346
HadCM3	2.575	2.402	2.111	0.602	NA	NA	111.246	2.404
inmcm4	2.623	2.453	2.101	0.381	2.358	2.063	111.153	2.270
FGOALS−s2	2.877	2.662	2.253	1.034	2.558	2.263	114.429	2.526
CSIRO−Mk3−6−0	2.961	2.763	2.393	0.977	2.445	2.127	115.992	2.612
FGOALS−g2	3.186	3.000	2.703	1.680	2.762	2.492	121.518	2.864
IPSL−CM5B−LR	3.361	3.238	2.929	0.626	2.741	2.421	117.348	2.731

Table 2.4: CMIP5 model rankings for (2m) surface temperature based on similarity to the ERA5 Reanalysis. Distances are calculated using our proposed spherical convolutional WD (SCWD) as well as the global mean-based WD (GMWD), RMSE and MAE (both computed on the climatologies), WD and Sliced WD (both computed on data regridded to an icosahedral grid). For the SCWD calculations, three different range parameters are chosen for the Wendland kernel: 500km, 1000km (our proposed choice), and 2500km. Color fill is unique to each column in the table, and is calculated using ranks.

CMIP5 Total Precipitation Rankings

	SCWD (500km)	SCWD (1000km)	SCWD (2500km)	GMWD	RMSE (climatology)	MAE (climatology)	WD (icosahedral)	SWD (icosahedral)
ERA5	1.804	1.133	0.701	0.252	1.936	1.349	155.514	0.928
ACCESS1−0	2.257	1.665	1.033	0.410	2.887	2.099	196.045	1.417
HadGEM2−ES	2.274	1.668	1.039	0.394	NA	NA	196.479	1.339
HadGEM2−CC	2.337	1.713	1.058	0.364	NA	NA	195.209	1.274
HadGEM2−AO	2.387	1.733	1.067	0.428	NA	NA	199.118	1.446
MIROC5	2.380	1.752	1.082	0.533	2.772	2.038	185.439	1.132
MIROC4h	2.809	1.817	1.080	0.278	2.937	2.053	204.359	1.765
EC−EARTH	2.334	1.859	1.156	0.199	2.581	1.879	171.102	1.462
ACCESS1−3	2.642	1.913	1.146	0.487	2.930	2.138	193.608	1.500
FGOALS−g2	2.774	1.913	1.109	0.132	2.969	2.104	202.659	1.566
NorESM1−M	2.435	1.937	1.212	0.147	2.725	1.959	172.984	1.463
CanESM2	2.767	1.989	1.185	0.107	2.719	1.961	179.564	1.251
CMCC−CMS	2.498	2.002	1.214	0.251	2.893	2.080	192.831	1.194
MPI−ESM−LR	2.527	2.009	1.169	0.275	2.896	2.091	190.268	1.292
GFDL−CM3	2.695	2.011	1.194	0.319	2.771	2.006	177.580	1.441
HadCM3	2.550	2.027	1.266	0.234	NA	NA	176.560	1.744
CanCM4	2.855	2.028	1.188	0.111	2.713	1.954	179.501	1.257
CMCC−CM	3.011	2.049	1.156	0.227	3.057	2.173	206.521	1.803
CMCC−CESM	2.685	2.129	1.319	0.204	2.798	2.045	176.287	1.653
MPI−ESM−MR	2.652	2.129	1.253	0.331	2.956	2.127	193.569	1.328
MPI−ESM−P	2.686	2.148	1.255	0.265	2.927	2.119	190.158	1.335
IPSL−CM5A−MR	2.914	2.189	1.315	0.144	2.787	2.030	177.914	1.576
MIROC−ESM	2.727	2.211	1.399	0.134	2.718	1.965	170.350	1.896
MIROC−ESM−CHEM	2.749	2.230	1.417	0.118	2.715	1.961	170.061	1.907
IPSL−CM5A−LR	2.882	2.230	1.333	0.080	2.709	1.968	169.559	1.867
GFDL−ESM2G	2.971	2.242	1.353	0.309	2.765	2.009	176.612	1.568
MRI−CGCM3	3.590	2.276	1.248	0.237	3.148	2.204	214.016	2.192
GFDL−ESM2M	3.280	2.279	1.232	0.317	2.772	2.005	178.057	1.386
MRI−ESM1	3.698	2.288	1.243	0.257	3.164	2.210	216.396	2.297
CSIRO−Mk3−6−0	2.898	2.326	1.404	0.213	2.913	2.136	182.964	1.520
FGOALS−s2	3.199	2.367	1.342	0.086	2.946	2.109	194.122	1.432
CNRM−CM5	3.764	2.412	1.218	0.384	2.828	2.039	189.153	1.254
NCEP	3.738	2.490	1.289	0.581	2.758	1.943	215.618	2.533
inmcm4	3.151	2.570	1.621	0.481	2.689	2.069	168.184	2.351
IPSL−CM5B−LR	4.504	2.630	1.168	0.145	3.228	2.196	219.657	2.561

Table 2.5: CMIP5 model rankings for total precipitation based on similarity to the GPCP observations. Distances are calculated using our proposed spherical convolutional WD (SCWD) as well as the global mean-based WD (GMWD), RMSE and MAE (both computed on the climatologies), WD and Sliced WD (both computed on data regridded to an icosahedral grid). For the SCWD calculations, three different range parameters are chosen for the Wendland kernel: 500km, 1000km (our proposed choice), and 2500km. Color fill is unique to each column in the table, and is calculated using ranks.

2.7.5 Investigation of Differences Between Rankings

The impacts of climate change on temperature are not limited to changes in the local means and may include changes in variance or other moments. This phenomenon is documented in Hansen et al. [2012], where extreme tail behavior in local temperature anomalies are shown to occur alongside warming in the mean climate state. We construct the following synthetic experiment to assess the ability of each climate model evaluation method (SCWD, GMWD, RMSE, and MAE) to detect changes in regional means and variances simultaneously. Based on these findings, we re-examine the CMIP6 surface temperature rankings in Table 2.2 and show that RSME/MAE rank SAM0-UNICON artificially high due to their inability to detect variance changes in the temperature anomalies.

Synthetic Experiment

To generate a realistic synthetic example, we start with the ERA5 data and linearly transform it to modify its mean and variance structure. Specifically, let Y be the ERA5 reanalysis surface temperature data. Y can be separated into the climatology, C, which is the temporal mean climate state for each location, and the anomalies $A = Y - C$. We compare $Y = C + A$ to modified datasets of the form:

$$Y_{M,s} = C + M + s * A$$

which are alternate versions of ERA5 with a mean shift of M and anomaly scale factor of s. We consider $s = 1.1$, $s = 1.3$, and $s = 1.5$, which respectively represent 10%, 30%, and 50% increases in the variance of the anomalies. We also consider three cases of M which vary in the Northern and Southern Hemispheres (NH and SH), including $(+0.5K, +0K)$ in NH/SH, $(+1K, -0.5K)$ in NH/SH, and $(+1.5K, -1K)$ in NH/SH. These are chosen to provide the same global mean change with different interhemispheric temperature asymmetries, which are an important climate feature [Friedman et al., 2013]. For each value of M and s, we compare $Y_{M,s}$ to Y using all four metrics.

The results are shown in Figure 2.5. Because each M has the same global mean, GMWD does not change with M, although it does (very slightly) increase for higher values of s. Conversely, s scales only the variance of the anomalies and does not impact the climatology, so RMSE/MAE are only able to detect changes in M, not s. Only SCWD detects changes in both the local means and the variance of the anomalies in this example.

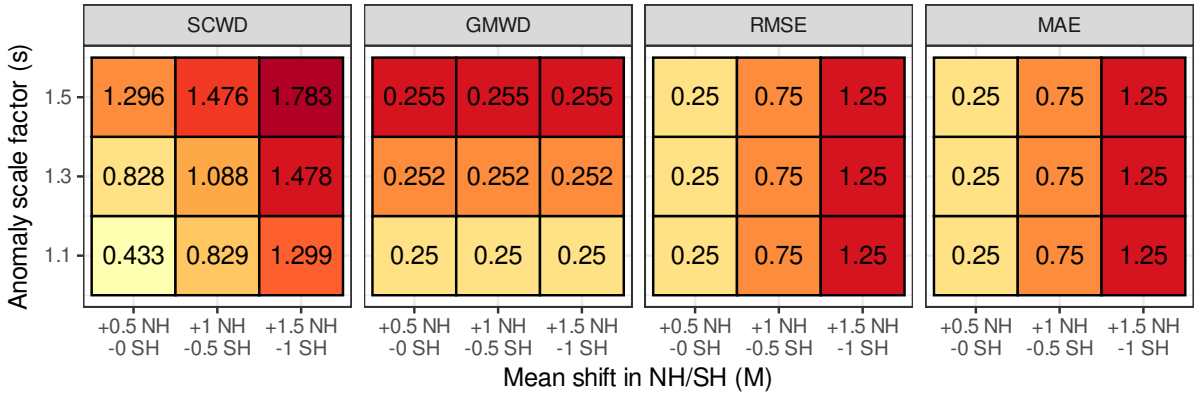


Figure 2.5: Results from the simulation comparing the original ERA5 data to synthetic modifications. Each panel gives the results for a different distance function. The y axis represents the anomaly scale parameter, s, and the x axis represents the M parameter which controls mean shift in the northern/southern hemispheres. For each method, the distance is provided from the original ERA5 data to the modified ERA5 data with each of the nine combinations of M and s. The color fill is determined by the rankings within each method, with light yellow representing a low ranking and dark red representing a high ranking.

Investigating Differences in CMIP6 Rankings

The experiment demonstrates that SCWD is able to distinguish additional sources of variability beyond just differences in the climatological mean state. To understand how this ability impacts our climate model rankings, we investigate SAM0-UNICON, which ranks poorly using SCWD relative to RMSE/MAE. Upon investigation of the local WD values, SAM0-UNICON exhibits the highest local WD values in regions such as the coast of Antarctica. We average over the region where the local WD is the highest and compare the climatologies and anomalies for ERA5 and SAM0-UNICON in Figure 2.6.

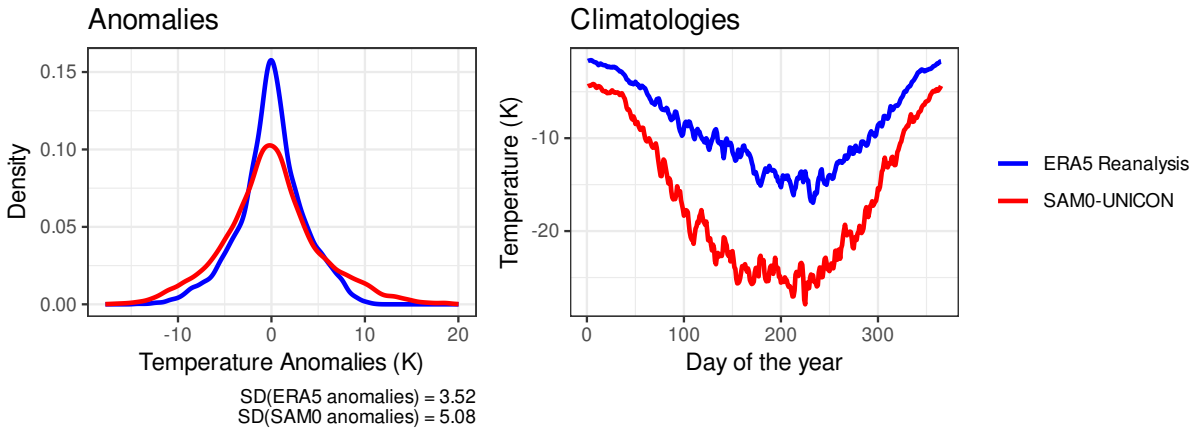


Figure 2.6: Comparison of anomalies and climatologies in the region where SAM0-UNICON has the largest local WD values (greater than 7.91). ERA5 is shown in blue and SAM0-UNICON is shown in red. The left plot is a density plot showing the distribution of the anomalies, and the right plot is a time series plot showing the climatologies starting on January 1st and ending on December 31st.

Relative to ERA5, we see a large difference in the climatology for SAM0-UNICON, which will impact both SCWD and RMSE/MAE. However, there is also a large increase in the variance of the anomalies in SAM0-UNICON, which is only measured by SCWD. We have demonstrated that this region has both high local WD values and large differences that are undetected by typical comparisons of the climatologies. So, this region is likely responsible in part for the harsher rankings of SAM0-UNICON by SCWD compared to RMSE/MAE.

Chapter 3

Sliced Elastic Distance for Evaluating Amplitude and Phase Differences in Precipitation Models

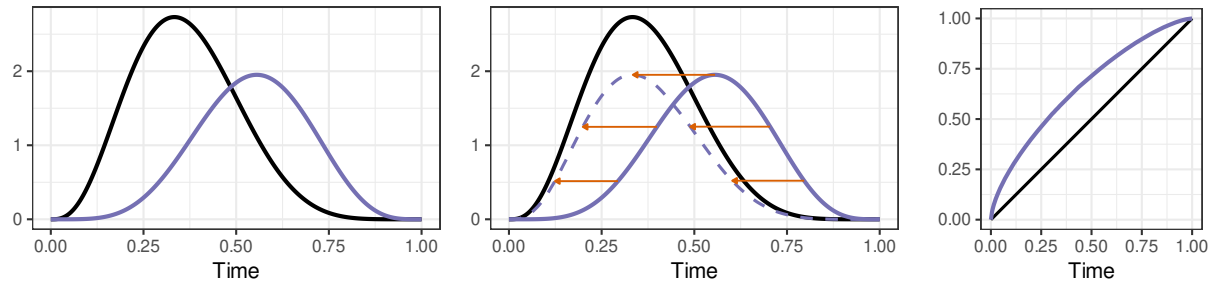
3.1 Introduction

[1]Climate models are an important tool to study the response of the climate system to antropogenic forcing and understand future variability [Kattenberg et al., 1996]. Precipitation is one of the most important variables affected by climate change due to its direct impact on agriculture, ecology, water management, and extreme weather events [Trenberth, 2011]. In particular, monsoon systems, which are characterized by prominent changes in rainfall and wind patterns, have broad impacts on agriculture and economics [Wang et al., 2006, Prasanna, 2014], renewable energy generation [Ramachandra and Shruthi, 2007], and extreme events such as floods and droughts [Bhalme and Mooley, 1980]. Due to these impacts, and because increases in extreme monsoon precipitation are associated with climate change [Zhang and Zhou, 2019], the accurate characterization of precipitation patterns is a crucial skill for climate models. Many new climate models introduced in the Coupled Model Intercomparison Project Phase 6 (CMIP6) present significant advances in physics parameterization and resolution aiming for better performance [Eyring et al., 2016b]. It is therefore essential to develop comprehensive climate model evaluation metrics to assess the skills of these models in capturing the complex dynamics of precipitation [Randall et al., 2007, Eyring et al., 2019].

[1]This chapter is based on the following manuscript in preparation: Garrett, R. C., Harris, T., Li, B., & Wang, Z. (2024). Sliced Elastic Distance for Evaluating Amplitude and Phase Differences in Precipitation Models.

Climate model evaluation centers around comparing model outputs to observational datasets [Flato et al., 2014]. Common similarity measures for comparing precipitation simulations to observations include root mean square error (RMSE), mean absolute error (MAE), Taylor skill score, interannual variability score, and correlation coefficients [Li et al., 2022, Yazdandoost et al., 2021, Du et al., 2022, Ngoma et al., 2021]. Evaluations of monsoon precipitation also rely on these similarity measures, and additionally focus on comparisons of seasonal mean precipitation during the monsoon event [Katzenberger et al., 2021, Konda and Vissa, 2023, Xin et al., 2020]. Since monsoon precipitation is important for agriculture, which is sensitive to both the timing and amount of rainfall, recent studies have focused on understanding biases in the onset and retreat dates of monsoon events [Ye and Wang, 2023, Khadka et al., 2022, Ha et al., 2020].

In the statistical literature, many methods have been developed to compare spatiotemporal fields which could be applied to the evaluation of precipitation models. These include comparisons based on the wavelet decomposition [Briggs and Levine, 1997, Shen et al., 2002, Cressie et al., 2008], loss differential [Snell et al., 2000, Wang et al., 2007, Hering and Genton, 2011], and first/second order dependency structure of spatial processes [Lund and Li, 2009, Li and Smerdon, 2012]. Other methods, such as [Yun et al., 2022], focus on identifying local differences in the spatial characteristics of climate fields. Functional data analysis methods can also be used to compare precipitation, including comparisons of the mean [Ramsay and Silverman, 2005, Zhang and Chen, 2007, Horváth et al., 2013, Staicu et al., 2014] or variance structure [Zhang and Shao, 2015, Li et al., 2016] of two collections of functions. Harris et al. [2021a] introduced a functional data depth method to nonparametrically compare the distribution of two spatial random processes.



(a) Original functions (b) Alignment Step (c) Warping function

Figure 3.1: Amplitude and phase separation using elastic functional data analysis. Plot (a) shows the two original functions. Plot (b) adds a dashed line to represent the purple curve after phase alignment to the black curve. The orange arrows represent the direction and severity of time warping. Plot (c) shows the time warping function in purple used in the phase alignment step.

Despite these recent advancements, existing methods have so far not considered the *shape* and *phase* differences between two spatiotemporal fields, which may be desirable for evaluating certain time sensitive climate patterns such as monsoon systems. Elastic functional data analysis fills in this void by decomposing

the total variation between two functions of time into amplitude (shape) distance, phase distance, and biases [Srivastava et al., 2011, Srivastava and Klassen, 2016]. A demonstration of amplitude and phase separation is given in Figure 3.1. Amplitude distance is calculated between two functions after time alignment and phase distance is calculated as the severity of time warping required for alignment. Because of the alignment step, amplitude differences can characterize deviations in the trajectory, or dynamical evolution, of the simulated climate compared with the observations, irrespective of mistiming of events. On the other hand, phase differences are focused solely on the timing of seasonal transitions or other weather events, which is of particular interest for the onset/retreat of seasonal monsoon events [Ye and Wang, 2023]. Elastic functional data analysis methods have been developed for various spatial data applications [Harris et al., 2021b, Guo et al., 2022, Tucker and Yarger, 2024], however so far there exists no adaptation of the amplitude and phase distances to compare functions with a joint space-time domain.

To create a tailored tool for precipitation model evaluation, we propose a sliced elastic distance metric that can separately quantify differences due to time misalignment from differences in the underlying dynamics of climate trajectories. Inspired by the sliced Wasserstein distance [Rabin et al., 2011], our approach computes amplitude, phase, and translation distances for many projections of climate fields, called slices. The slices are computed using a kernel convolution approach to isolate local spatial features of interest. Compared to previous approaches, the sliced elastic distance accounts for both the location and timing of precipitation events when comparing climate models against observational data, resulting in a more comprehensive evaluation. Additionally, our metric can provide shape and phase differences at every location, offering useful insights to climate scientists. We apply our method to rank CMIP6 daily precipitation models against observational data and map local summaries of the onset and retreat bias of the Indian Summer Monsoon in each model.

The rest of the article is organized as follows. Section 3.2 describes the precipitation data for our analysis. Section 3.3 proposes the sliced elastic distance and establishes its theoretical properties. Section 3.4 demonstrates the performance of the sliced elastic distance on simulated model validation scenarios. Section 3.5 applies our distance to evaluate the CMIP6 precipitation models in terms of shape and phase differences on both a global and local scale. Section 3.6 provides a brief discussion of the method and the results.

3.2 Data Description

We obtain precipitation outputs from the Coupled Model Intercomparison Project Phase 6 (CMIP6) historical experiment, available at `https://esgf-node.llnl.gov/projects/esgf-llnl/`, to assess the global and local performance of climate models. The total precipitation (in mm) is obtained at a daily frequency, allowing us to study the seasonal transitions and trends in climate model simulations. Ensembles from the

CMIP6 historical simulation are distinguished with `ripf` identifiers, representing each model's realization, initialization, physics, and forcings [Eyring et al., 2016b]. We obtain 45 model outputs with the `r1i1p1f1` variant ID, each representing a single simulation of the historical precipitation field from a different model. See Table 3.1 in Section 3.8.3 for a full list of models included in the analysis.

We collect the National Centers for Environmental Information (NCEI) Global Precipitation Climatology Project (GPCP) Daily Precipitation Analysis Climate Data Record [Huffman et al., 2001, Adler et al., 2020] as the reference for climate model evaluation. Data were obtained from the GPCP Daily V1.3 analysis and are publicly available on the NCEI website: https://www.ncei.noaa.gov/products/climate-data-records/precipitation-gpcp-daily. To establish the baseline similarity between reanalysis datasets and GPCP relative to CMIP models and GPCP, we also include precipitation fields from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis 5th Generation (ERA5) [Hersbach et al., 2020, 2023] and the National Centers for Environmental Protection (NCEP) Reanalysis-2 dataset [Kanamitsu et al., 2002]. The ERA5 hourly data are available at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview and were subsequently aggregated to a daily frequency and the NCEP Reanalysis-2 daily data are available at https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html.

Each data product provides precipitation values on a rectangular grid of latitude-longitude coordinates across the Earth except for a single climate model, ICON-ESM-LR. The spatial resolution of the grid varies for different data products. Model evaluation is conducted during the historical period of January 1997 to December 2014, a common period covered by GPCP, ERA5, NCEP, and the 45 CMIP6 models represented in our study.

3.3 Methods

3.3.1 Review of elastic functional data analysis

Elastic functional data analysis (EFDA) [Joshi et al., 2007, Tucker et al., 2013, Srivastava and Klassen, 2016] is a framework for comparing the shapes of absolutely continuous manifold-valued functions parameterized by time. To serve our application to phase comparison of precipitation data, we consider a special case of functions in the space $\mathcal{F}_{\mathcal{T}}$, the set of absolutely continuous functions from $\mathcal{T} \mapsto \mathbb{R}$. The time domain $\mathcal{T}$ is assumed to be $[0, 1]$ without loss of generality. To compare the shapes of two functions $f(t), g(t) \in \mathcal{F}_{\mathcal{T}}$, EFDA uses time warping to decompose the variability between f and g into amplitude, phase, and translation distances. On the space $\mathcal{F}_{\mathcal{T}}$, each of the three distances alone are not valid distance metrics because they do

not individually satisfy the positivity property of a metric: for all $f, g \in \mathcal{F}_{\mathcal{T}}$, $f \neq g \Rightarrow D(f,g) > 0$, where $D(f,g)$ denotes a distance between f and g.

To create proper amplitude, phase, and translation metrics, EFDA introduces the square root velocity function (SRVF) representation [Joshi et al., 2007] along with time warping to decompose functional data into three component spaces. The SRVF of a function $f \in \mathcal{F}_{\mathcal{T}}$ is defined as $q_f(t) = \text{sign}(\dot{f}(t))\sqrt{|\dot{f}(t)|}$ [Srivastava and Klassen, 2016], where $\dot{f}(t)$ is the derivative of $f(t)$ with respect to time t. Because q_f is an element of $\mathbb{L}_2$, the set of square-integrable functions, it is a unique and invertible representation of f up to vertical translation, represented by the intercept $f(0)$. We then use $\gamma_f(t) \in \Gamma$ to denote a warping function for q_f, where Γ is the set of functions on $\mathcal{T}$ which are boundary preserving, absolutely continuous, and weakly increasing. Let D_A, D_P, and D_T denote the valid amplitude, phase, and translation distance metrics between f and g, respectively. Srivastava and Klassen [2016] defines these distances as follows:

$$\begin{aligned} D_A(f,g) &= \inf_{\gamma_f, \gamma_g \in \Gamma} ||(q_f, \gamma_f) - (q_g, \gamma_g)||_2, \\ D_P(f,g) &= \cos^{-1}\left(\int_0^1 \sqrt{\dot{\gamma}_f^*(t)}\sqrt{\dot{\gamma}_g^*(t)}dt\right), \\ D_T(f,g) &= |f(0) - g(0)|, \end{aligned} \tag{3.1}$$

where $(q_f, \gamma) = (q_f(\gamma(t)))\sqrt{\dot{\gamma}}$ is the SRVF of $f(\gamma(t))$, the time warping of f by γ, and $\dot{\gamma}$ is the derivative of γ with respect to time. Phase distance between f and g is defined as a function of their relative phase functions, denoted by γ_f^* and γ_g^*, which are the minimizers obtained in the calculation of $D_A(f,g)$. The values $f(0)$ and $g(0)$ are the time intercepts of each function. The distance functions D_A, D_P, and D_T are all proper metrics on a different space from the SRVF representation [Srivastava and Klassen, 2016]. Amplitude distance is a metric on the quotient space $\mathbb{L}_2/\Gamma$ and provides a comparison of q_f and q_g after accounting for phase variability. Phase distance is a metric on Γ, and quantifies the severity of the time warping required to align q_f and q_g. Translation distance is a metric on $\mathbb{R}$ and represents the magnitude of the difference between the intercepts of each function.

The triplet of amplitude, phase, and translation distance together is called the elastic distance, and is denoted as the following vector:

$$D_E(f,g) = \Big[D_A(f,g), D_P(f,g), D_T(f,g)\Big]^T, \tag{3.2}$$

where x^T refers to the transpose of a row vector x. EFDA centers on identifying the elastic distance between functions. Even though D_A, D_P, and D_T are metrics on different spaces, when considered jointly, they fully characterize the differences between two functions. In fact, we prove that D_E is a vector-valued metric [Sastry

et al., 2012] on the space $\mathcal{F}_{\mathcal{T}}$ in Section 3.3.3.

3.3.2 Sliced elastic distance

Spatiotemporal climate fields can be viewed as functional data indexed by spatial location s and time t, denoted by $f(s,t) \in \mathcal{F}_{\mathcal{S}\times\mathcal{T}}$, where $\mathcal{F}_{\mathcal{S}\times\mathcal{T}}$ is the set of continuous functions from $\mathcal{S}\times\mathcal{T} \mapsto \mathbb{R}$. The spatial domain $\mathcal{S}$ is the Earth's surface, which is indexed by latitude and longitude coordinates from the unit sphere $\mathbb{S}^2$. The time domain $\mathcal{T}$ can be any time interval of interest, and is represented by the interval $[0,1]$. To meet typical EFDA assumptions, we additionally assume that for all $f \in \mathcal{F}_{\mathcal{S}\times\mathcal{T}}$, if the location is fixed at any $s' \in \mathcal{S}$, then $f(s',t)$ is an absolutely continuous function of time. To simplify notation, we will suppress (s,t) from $f(s,t)$ unless there is a risk of confusion.

Although such data can be considered as multivariate functions of time so that traditional EFDA methods may apply [Joshi et al., 2007, Srivastava and Klassen, 2016], there are two main issues with this approach. Firstly, traditional EFDA necessitates that the functions being compared have a shared set of p dimensions, translating to a requirement that each climate model output and reanalysis product share a common spatial grid. However, in practice, the size and structure of the spatial grid often varies between data products. Secondly, traditional EFDA does not account for spatial variability and allows only a single warping function to align two $p-$variate functional data [Srivastava and Klassen, 2016, Joshi et al., 2007, Bernal et al., 2021, Tucker et al., 2022, Hartman et al., 2021]. Whereas, climate fields exhibit spatially-varying phase characteristics such as seasonal transitions varying by latitude as well as alternating in the northern and southern hemispheres. Our framework instead treats $\mathcal{S}$ as part of the function domain, naturally integrating spatial information into the data. This enables our method to combine data from nearby locations while simultaneously allowing time warping to vary at different locations, thereby accounting for phase variability in different regions of the globe.

Similar to the sliced Wasserstein distance [Rabin et al., 2011] proposed for high-dimensional distributions, we introduce a sliced elastic distance to compare two spatiotemporal fields. Slices are typically constructed as low-dimensional projections of the data. Considering the various grid sizes and structures of different climate data products, we propose to construct slices at arbitrary locations by convolving nearby functional data using a kernel function centered at each slice location. Specifically, we define the slice function, $f_s(t)$, of $f(u,t) \in \mathcal{F}_{S\times T}$ at location $s \in \mathcal{S}$ as

$$f_s(t) = \int_{\mathcal{S}} f(u,t) k_s(u;\theta) du, \tag{3.3}$$

where $k_s(u;\theta)$ is a spatial kernel function with parameter θ centered at location s. The kernel can be any spatially continuous function with a positive spectral density on $\mathcal{S}$. Analogously, we can define the slice

function $g_s(t)$ for $g(u,t) \in \mathcal{F}_{S\times T}$. It is easy to compute $D_E(f_s, g_s)$ as both f_s, g_s are univariate functional data. We then calculate the sliced elastic distance between the two spatiotemporal fields as the global mean of the elastic distances between the slice functions at each location in $\mathcal{S}$.

Definition 3.1 (Sliced elastic distance)**.** Let $f, g \in \mathcal{F}_{\mathcal{S}\times\mathcal{T}}$. We define the sliced elastic distance, D_{SE}, between f and g to be the vector consisting of three components: sliced amplitude distance (D_{SA}), sliced phase distance (D_{SP}), and sliced translation distance (D_{ST}):

$$D_{SE}(f,g) = \begin{bmatrix} D_{SA}(f,g) \\ D_{SP}(f,g) \\ D_{ST}(f,g) \end{bmatrix} = \begin{bmatrix} \left\{\int_{\mathcal{S}} D_A(f_s,g_s)^2 ds\right\}^{1/2} \\ \left\{\int_{\mathcal{S}} D_P(f_s,g_s)^2 ds\right\}^{1/2} \\ \left\{\int_{\mathcal{S}} D_T(f_s,g_s)^2 ds\right\}^{1/2} \end{bmatrix}, \tag{3.4}$$

where f_s and g_s are the slices defined in (3.3) and D_A, D_P, and D_T are amplitude, phase, and translation distances between two univariate functions as defined in (3.1).

Since D_{SA} is calculated via the amplitude distance between slices f_s and g_s at each location $s \in \mathcal{S}$, the time warping step is allowed to vary over space. According to (3.1), the amplitude distance is based on the derivative of each slice. Thus, D_{SA} integrates local differences between the dynamics of two climate fields in the vertical direction. The phase distance between f_s and g_s represents the severity of the warping required to align the SRVFs of f_s and g_s. When comparing climate fields, $D_P(f_s, g_s)$ quantifies differences in the timing of weather events and seasonal changes at location s, and $D_{SP}(f,g)$ measures the average time misalignment between the two climate fields. Sliced translation distance only captures differences between two functions at $t = 0$. Though it is a necessary component to ensure that the sliced elastic distance satisfies the properties of a vector-valued metric, it is of little scientific interest for climate field comparison. We recommend to substitute the sliced translation distance by a different measure of bias that is more of interest, such as bias in the annual mean or over a particular season of interest. We focus on D_{SA} and D_{SP} in our study.

The sliced elastic distance allows us to compare high-dimensional spatiotemporal data through one-dimensional slices indexed by spatial locations. Each slice represents one perspective of the data with amplitude and phase distances calculated using the usual dynamic programming algorithm [Joshi et al., 2007], therefore avoiding computational concerns with the time alignment step for multivariate functional data [Bernal et al., 2021, Tucker et al., 2022, Hartman et al., 2021]. Together, the slices provide a comprehensive view of a spatiotemporal field. Figure 3.2 illustrates the idea of climate fields comparison using sliced elastic distance. The kernel convolution approach allows for the comparison of climate model output and observational data that are available at different spatial resolutions. Details for computing the sliced elastic distance are provided in Section 3.8.2.

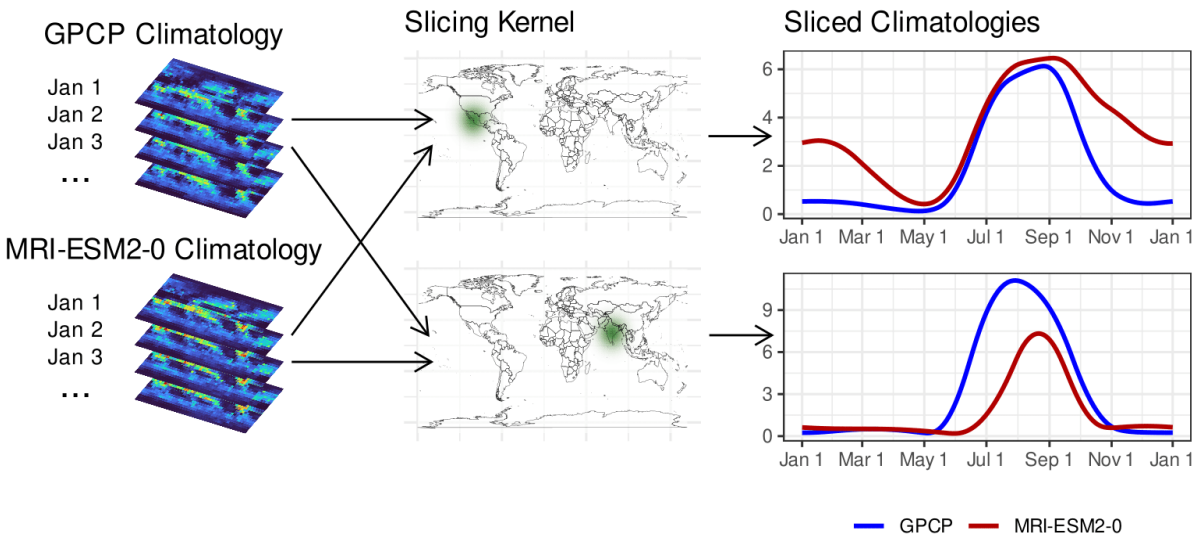


Figure 3.2: Constructing the slice functions for GPCP and one CMIP6 model. Each day, the spatial fields from each dataset are converted into many univariate functions, or slice functions, using a set of kernel projections. The slice functions are used to compare each model to GPCP via the elastic distance metric (3.1). The sliced elastic distance is obtained by averaging the elastic distance for each pair of slice functions.

Choice of kernel function

Valid choices of kernel functions include the Kent distribution function [Kent, 1982] and the generalized Wendland functions [Wendland, 1998]. Compact kernels are desirable for our application because they ensure each slice represents climate features in a relatively small local neighborhood, though ultimately our method provides a global characterization of the misalignment between two fields. We thus choose the Wendland kernel function as in Nychka et al. [2015]:

$$k_s(u;r) = \begin{cases} \left(1 - \frac{|s-u|}{r}\right)^6 \left(35\frac{|s-u|^2}{r^2} + 18\frac{|s-u|}{r} + 3\right)/3 & |s-u| \le r, \\ 0 & |s-u| > r, \end{cases} \tag{3.5}$$

where the range parameter r determines the compactness of the kernel. In our application, $|s-u|$ is the chordal distance between locations s and u, represented as latitude-longitude coordinates. Chordal distance is chosen for theoretical convenience. Suppose the range parameter is chosen to be smaller than or equal to the diameter of the Earth. Then the Wendland functions will be positive definite on $\mathbb{S}^2$ [Hubbert and Jäger, 2023], therefore meeting the assumptions in Definition 3.1.

Timing Biases

The sliced elastic distance between two functions $f(u,t)$ and $g(u,t)$ is calculated as a global mean of the elastic distances between the localized sliced functions f_s and $g_s, s \in \mathcal{S}$. Calculating the local elastic distance between each pair of sliced functions produces one useful intermediate output, the relative phase functions $\gamma^*_{f_s}$ and $\gamma^*_{g_s}$. As discussed in Section 3.8.2, the code implementation for calculating D_A and D_P assumes that $\gamma^*_{f_s} = I(t)$ for identifiability, where $I(t) = t, t \in \mathcal{T}$ is the identity warping function [Tucker et al., 2013]. Under this assumption, no time warping is applied to f_s, thus $\gamma^*_{g_s}(t)$ represents the optimal warping function to align g_s to f_s. Therefore, for a given time $t \in \mathcal{T}$, we can calculate the timing bias, denoted as $B(f_s, g_s; t)$, of g_s relative to f_s as follows:

$$B(f_s, g_s; t) = \gamma^*_{g_s}(t) - I(t). \tag{3.6}$$

In contrast to phase distance, timing bias is focused around a single user-specified time point. Additionally, timing bias can be positive or negative, representing late or early timing of an event in g_s compared to f_s, respectively. In our application, f is fixed to be the observed precipitation data, and g represents any climate model data. This allows us to characterize timing biases for events of interest for each climate model relative to the observed data.

3.3.3 Theoretical properties

We first show that the elastic distance $D_E(f(t), g(t))$ defined in (3.1) is a valid vector-valued metric on $\mathcal{F}_\mathcal{T}$ and then establish our main result that the sliced elastic distance $D_{SE}(f(s,t), g(s,t))$ is a valid vector-valued metric on $\mathcal{F}_{\mathcal{S}\times\mathcal{T}}$. Both results rely on the idea of vector-valued metric spaces [Sastry et al., 2012, Rao, 2015, Jachymski and Klima, 2016] defined below, which generalizes the concept of metric spaces to allow for multiple distance functions.

Definition 3.2 (Vector-valued metric)**.** Let X be a nonempty set, and let $D : X \times X \to \mathbb{R}^m$ be a length m vector-valued function with the following properties:

1. Identity: $D(x,x) = 0_m$ for all $x \in X$,
2. Symmetry: $D(x,y) = D(y,x)$ for all $x, y \in X$,
3. Triangle Inequality: $D(x,y) \leq D(x,z) + D(z,y)$ for all $x, y, z \in X$,
4. Positivity: $0_m \leq D(x,y)$ for all $x, y \in X$ such that $x \neq y$,

where 0_m is the zero vector of length m, and $a \leq b$ for $a, b \in \mathbb{R}^m$ holds if and only if $a_i \leq b_i$ for all $i \in 1, ..., m$. We say D is a vector-valued metric with m components and (X, D) is a vector-valued metric space.

Jachymski and Klima [2016] introduced an alternate characterization of vector-valued metrics which states that (X, D) is a vector-valued metric space if and only if D is a family of pseudometrics $(D_1, ..., D_m)$ such that for any $x, y \in X$, $x \neq y \Rightarrow D_i(x, y) > 0$ for some $i \in 1, ..., m$. This characterization fits naturally with the EFDA representation. Section 3.3.1 reviews the three components of EFDA: amplitude, phase, and translation distance. Each is a valid metric on a different space, but not on the original functional data space, $\mathcal{F}_{\mathcal{T}}$. However, it is obvious that if two functions $f(t)$ and $g(t)$ are not the same, there will be a disparity in either their aligned SRVFs, their relative phase functions, or their intercepts, leading to a positive value for the respective amplitude, phase, or translation distance. We use the characterization of vector-valued metrics from Jachymski and Klima [2016] along with properties from Srivastava and Klassen [2016] to prove the following Lemma.

Lemma 3.1. *$D_E(f(t), g(t))$ is a vector-valued metric on $\mathcal{F}_{\mathcal{T}}$.*

The proof of Lemma 3.1 is deferred to Section 3.8.1. Based on this result, we show that the sliced elastic distance D_{SE} is a vector-valued metric on $\mathcal{F}_{\mathcal{S}\times\mathcal{T}}$.

Theorem 3.1. *If $D = [D_1, ..., D_m]^T$ is a vector-valued metric on $\mathcal{F}_{\mathcal{T}}$, and $f_s(t)$ and $g_s(t)$ are respectively the slice functions of $f(u,t) \in \mathcal{F}_{\mathcal{S}\times\mathcal{T}}$ and $g(u,t) \in \mathcal{F}_{\mathcal{S}\times\mathcal{T}}$ using a spatially continuous kernel $k(u;\theta)$ with a positive spectral density on spherical domain $\mathcal{S} \in \mathbb{S}^2$ as defined in (3.3), then the vector-valued function $D_S = [D_{S1}, ..., D_{Sm}]^T$ with each component defined as*

$$D_{Si}(f, g) = \left\{ \int_{\mathcal{S}} D_i \left(f_s, g_s\right)^2 ds \right\}^{1/2}, \quad i = 1, ..., m,$$

is a vector-valued metric on $\mathcal{F}_{\mathcal{S}\times\mathcal{T}}$.

Corollary 3.1. *The sliced elastic distance $D_{SE}(f(s,t), g(s,t))$ is a vector-valued metric on $\mathcal{F}_{\mathcal{S}\times\mathcal{T}}$.*

Proof of Theorem 3.1 is provided in Section 3.8.1. This theorem shows that our slicing process extends the properties of any metric or vector-valued metric on $\mathcal{F}_{\mathcal{T}}$ to the space $\mathcal{F}_{\mathcal{S}\times\mathcal{T}}$. The proof relies on the convolution theorem for finite-dimensional unit spheres [Driscoll and Healy, 1994] as well as Lemmas 3.2 and 3.3. Lemma 3.2 states that the slicing operation in (3.3) creates valid functional data in $\mathcal{F}_{\mathcal{T}}$ and Lemma 3.3 states that pseudometrics on $\mathcal{F}_{\mathcal{T}}$ extend to pseudometrics on $\mathcal{F}_{\mathcal{S}\times\mathcal{T}}$ when our slicing method is applied. Corollary 3.1 follows directly from Lemma 3.1 and Theorem 3.1. Since the sliced elastic distance is a vector-valued metric on $\mathcal{F}_{\mathcal{S}\times\mathcal{T}}$, its components, the sliced amplitude, phase, and translation distances, extend the elastic distance to provide a comprehensive suite of metrics for comparing spatiotemporal fields. While our theoretical results are shown for spatial data with a spherical domain, similar results can be proven for any spatial domain on which a comparable convolution theorem is available.

3.4 Simulation

We conduct a numerical experiment to evaluate the skill of our method in separating amplitude and phase differences in spatiotemporal fields, investigate the sensitivity of our method to the choice of range parameter, and compare our method to a traditional precipitation evaluation method.

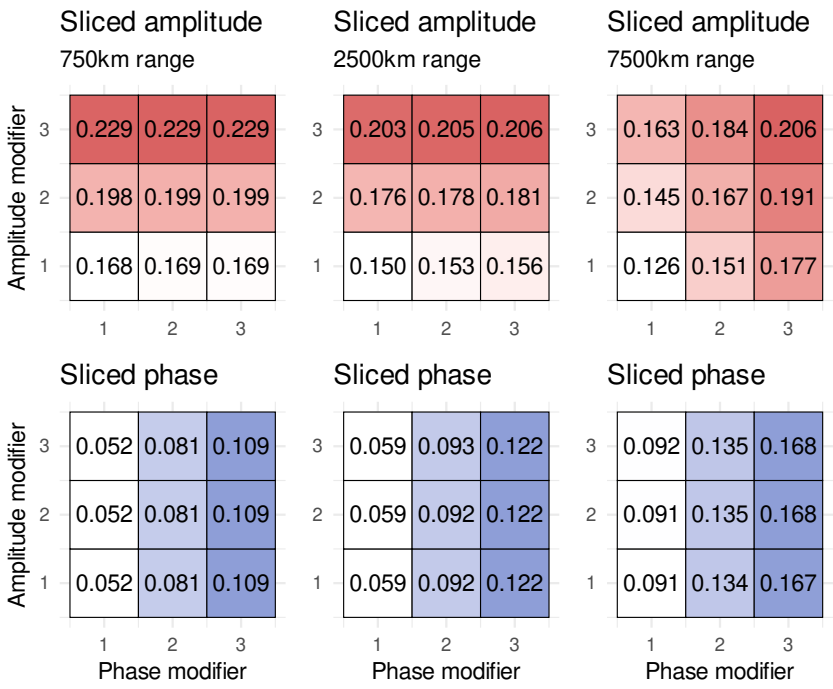
To make our simulation realistic, we convert the GPCP precipitation observations to climatologies by computing the mean over 1997-2014 at each location on each day of the year, excluding leap days. These GPCP climatologies serve as the baseline data $f(s,t)$. We then generate $g(s,t)$ by applying a series of spatially-varying amplitude and phase modifications on $f(s,t)$. Let $a_i(s)$, $i = 1,2,3$ and $p_j(s)$, $j = 1,2,3$ be the sets of parameters used for the amplitude and phase modifications, and let $g_{i,j}(s,t)$ denote the modified versions of GPCP corresponding to parameters $a_i(s)$ and $p_j(s)$. We obtain $g_{i,j}(s,t)$ by transforming $f(s,t)$ for all $s \in \mathbb{S}^2$ as follows

$$g_{i,j}(s,t) = a_i(s) f\left(s, t^{p_j(s)}\right). \tag{3.7}$$

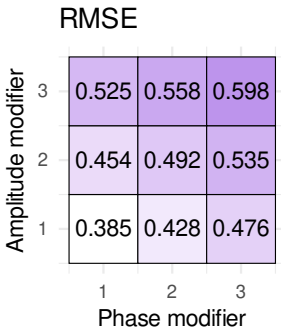
The amplitude parameter $a_i(s)$ acts as a multiplier on the scale of $f(s,t)$. We restrict $a_i(s)$ to values larger than 1 so that larger values of $a_i(s)$ correspond to increased levels of precipitation relative to the original $f(s,t)$, increasing the amplitude distance. The phase parameter $p_j(s)$, which we restrict to positive values, introduces time warping to change the seasonal timing of $f(s,t)$ at each loation. Values larger than 1 introduce late timing biases and values between 0 and 1 introduce early timing biases.

The parameter fields $a_i(s)$ and $p_j(s)$ are both generated as functions of latitude. The amplitude parameters, $a_i(s)$, all start at a value of 1.1 at the south pole and then linearly increase to 1.15, 1.2, and 1.25 at the north pole for $i = 1, 2$, and 3, respectively. This modification magnifies amplitude everywhere and intensifies with latitude. The phase parameters, $p_j(s)$, vary exponentially with latitude according to the functions $1.2^{\text{lat}(s)/90}$, $1.4^{\text{lat}(s)/90}$, and $1.6^{\text{lat}(s)/90}$ for $j = 1, 2$, and 3, respectively, where $\text{lat}(s)$ is the latitude coordinate of location s in degrees. In the southern hemisphere, the latitude coordinates are negative leading to $p_j(s) < 1$, so early timing biases are introduced in that region. On the contrary, the modifications introduce late timing biases in the northern hemisphere. For both $a_i(s)$ and $p_j(s)$, the most extreme amplitude and phase modifications occur when $i = j = 3$ and the least extreme modifications occur when $i = j = 1$.

We calculate the sliced amplitude and sliced phase distance between $f(s,t)$ and each $g_{i,j}(s,t)$ following the algorithm described in Section 3.8.2. To understand the influence of range parameter r in the Wendland kernel function $k(u;r)$ on the distances, we repeat the calculation for three different r values in km: 750, 2500, and 7500. The first value, $r = 750$ represents the value used in Section 3.5. The remaining values represent potential choices for larger kernel ranges, and each value is less than the Earth's diameter (approximately 12,742km) to ensure positive definiteness.



(a) Sliced Elastic Distance Tables



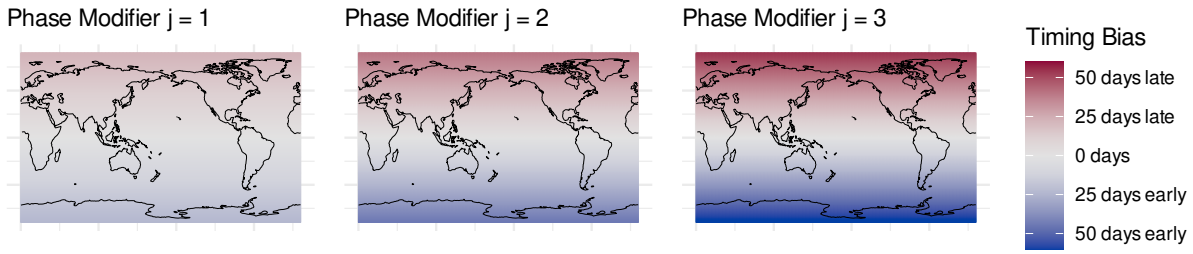
(b) RMSE Table

Figure 3.3: Panel (a): the sliced amplitude distance (red) and sliced phase distance (blue) between the original GPCP precipitation climatologies, $f(s,t)$, and the modified versions, $g_{i,j}(s,t)$, at various kernel range parameter values. Panel (b): the RMSE (purple) between the original and modified GPCP climatologies. For all plots, levels of the amplitude modifier ($i = 1, 2, 3$) and phase modifier ($j = 1, 2, 3$) are labeled in the x- and y-axis, respectively. Larger values of the amplitude and phase modifiers correspond to larger modifications made to $f(s,t)$. Color fill is determined independently per table, with lighter shades representing low distances and darker shades representing high distances.

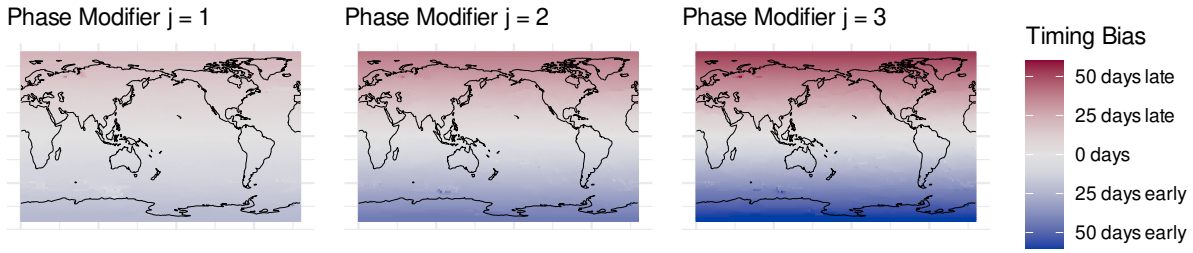
Simulation results are reported in Figure 3.3. Overall, the sliced amplitude and sliced phase distance patterns show that our method is able to separate the spatially-varying amplitude differences from the spatially-varying phase differences. However, the range parameter values can affect this ability. For the range value of 750km, there is little to no influence of the phase modifier $p_j(s)$ on the sliced amplitude distance, evidenced by the very consistent amplitude distances at the three different phase modifier values. Vice versa, we also see no influence of the amplitude parameter $a_i(s)$ on the sliced phase distance values. Similar patterns are observed even if the range value increases to 2500km. Whereas, for the range parameter of 7500km, the sliced amplitude distances increase with $p_j(s)$, showing evidence of entangled sliced amplitude and sliced phase distances. This is because data that are further away in space have more distinct phase characteristics in our simulation. Ignoring phase variability when taking functional means is known to distort the underlying structure [Tucker et al., 2013]. So, when using larger range parameter values, the kernel convolution used to create the slices acts as a cross-sectional weighted mean of misaligned functional data,

leading to the entangled distances. To more accurately quantify the amplitude and phase distances, we recommend choosing smaller range parameter values to decrease the influence of phase variability within the kernel radius. However, the grid size of data products must be considered in the choice of range parameter. If the range parameter is chosen to be significantly smaller than the grid size of the data, the slice functions may not contain the spatial information at the desired level.

Figure 3.3 (b) shows the RMSE calculated as the Euclidean norm between the original and modified climatologies. Unlike the sliced elastic distance, RMSE is unable to distinguish amplitude variability from phase variability, so RMSE provides distances that are essentially a joint reflection of the amplitude and phase modifications.



(a) True timing biases in the simulated data



(b) Timing biases estimated by our method

Figure 3.4: Maps of timing bias on July 2 in the simulated data example for the three levels of the phase modifier $p_j(s), j = 1, 2, 3$. Panel (a) shows the true timing biases at each location in the modified GPCP climatologies. Panel (b) shows the timing biases estimated using our sliced elastic distance method. Slices are computed using the 750km kernel, and timing biases are calculated using the formula in (3.6). The color fill represents the magnitude and direction of timing biases.

In addition to quantifying the overall phase distance between two spatiotemporal fields, our method can also provide detailed information about timing biases at a given time point. We compare the true timing biases in the modified GPCP climatologies incurred by the modifications in (3.7) to the values estimated by our method using the formula in (3.6). Figure 3.4 (a) shows maps of the true timing biases on July 2 (the midpoint of the calendar year), while Figure 3.4 (b) shows maps of the estimated values using the 750km kernel. These maps offer both the value and direction (early/late) of timing biases at each location for all three levels of phase modifications. In all cases, the fields are near-identical between the true and estimated

values. The correct magnitude and direction of timing biases are recovered, with late biases near the north pole, early biases near the south pole, and no bias near the equator.

3.5 Climate Model Evaluation

We first evaluate the skill of the CMIP6 climate models in reconstructing the historical climatologies on a global scale. We compute the sliced elastic distance between the daily precipitation fields of each CMIP6 model output and the GPCP observations. We repeat this process also for the ERA5 and NCEP datasets to evaluate the performance of reanalysis fields against GPCP, thus providing a baseline for climate model evaluation. To show the utility of our method in understanding timing variability in climate models, we use the intermediate results from the sliced elastic distance calculation to quantify the timing bias for the onset and retreat of the Indian Summer Monsoon. We first demonstrate this process at a single location, then repeat the procedure across the entire monsoon region for a cohort of six CMIP6 models to produce maps of timing biases relative to GPCP.

Before computing the distance, raw precipitation values from each dataset are converted into daily climatology fields by taking the mean over 1997-2014 at each location on each day of the year, excluding leap days. Climate model evaluation studies typically focus on monthly data rather than daily to provide data smoothing. However, the temporal resolution of monthly data is too coarse to accurately determine phase variability and timing biases, so we opt for daily data. Even after taking averages over many years, daily precipitation data can be rather noisy. Given our focus on the temporal evolution of precipitation, quadratic trend filtering [Tibshirani, 2011], which is designed to estimate the underlying continuous trend from noisy data, is applied to estimate a continuous function at each location in the climatology. This serves to reduce the noise present in daily observations while retaining detailed temporal information that is not captured in monthly mean data. For all sliced elastic distance computations, we follow the steps detailed in Section 3.8.2, and use the Wendland kernel function in (3.5) with range r =750km. This range is large enough to cover more than a three grid cell distance near the equator in the lowest-resolution CMIP6 model, ensuring that some degree of spatial smoothing is applied to all models. On the other hand, this choice is small enough to avoid issues with large range parameter values where sliced amplitude and sliced phase distances become entangled as demonstrated in Section 3.4.

3.5.1 Global evaluation of CMIP6 Precipitation Models

We create a 46-member ensemble of historical CMIP6 model outputs for daily total precipitation from January 1997 through December 2014, and compute the sliced elastic distance between the climatologies for each

model and the GPCP data during this time period. A smaller sliced amplitude distance indicates that the model more closely matches the shape of the observations, while a smaller sliced phase distance indicates closer agreement in the timing of events and seasons. As mentioned earlier, we also compute the distance from the ERA5 and NCEP Reanalysis fields to GPCP.

Figure 3.5 shows that ERA5 has lower sliced amplitude and phase distances to GPCP than all CMIP6 models and NCEP, indicating a higher degree of similarity to the observed data. NCEP has a lower sliced phase distance to GPCP than all CMIP6 models. However, two models, NorESM2-MM and CESM2-WACCM, have a lower sliced amplitude distance than NCEP to GPCP. Despite known issues with reanalysis datasets for precipitation [Tapiador et al., 2017], our method shows greater agreement between reanalysis datasets and observations than model outputs and observations, particularly in sliced phase distance.

Among the CMIP6 models, the scatter plot shows evidence of a positive correlation between sliced amplitude and sliced phase distance. This implies that a model which performs well in one component (amplitude or phase) likely performs well in the other. The Norwegian Earth System Model (NorESM2-MM) from the Norwegian Climate Center seems to rank the best with the lowest sliced amplitude distance by a wide margin and the second-lowest sliced phase distance. The Community Earth System Model (CESM) Whole Atmosphere Community Climate Model configuration (CESM2-WACCM) has the second-lowest sliced amplitude distance and the Alfred Wegener Institute Climate Model (AWI-CM-1-1-MR) has the lowest sliced-phase distance by a small margin. The Energy Exascale Earth System Models (E3SM) also show a favorable balance of sliced amplitude and phase distances. The Icosahedral Non-hydrostatic Earth System Model (ICON-ESM-LR) stands out the highest values for both sliced amplitude and sliced phase distance, while the Nanjing University of Information Science and Technology Earth System Model (NESM3) and the interactive aerosols/atmospheric chemistry configuration of the Institut Pierre-Simon Laplace coupled climate model (IPSL-CM5A2-INCA) have the second-highest sliced amplitude and sliced phase distances, respectively. For some climate modeling groups, such as E3SM and EC-Earth, there are strong similarities between each model in the group in terms of sliced amplitude and phase distance values.

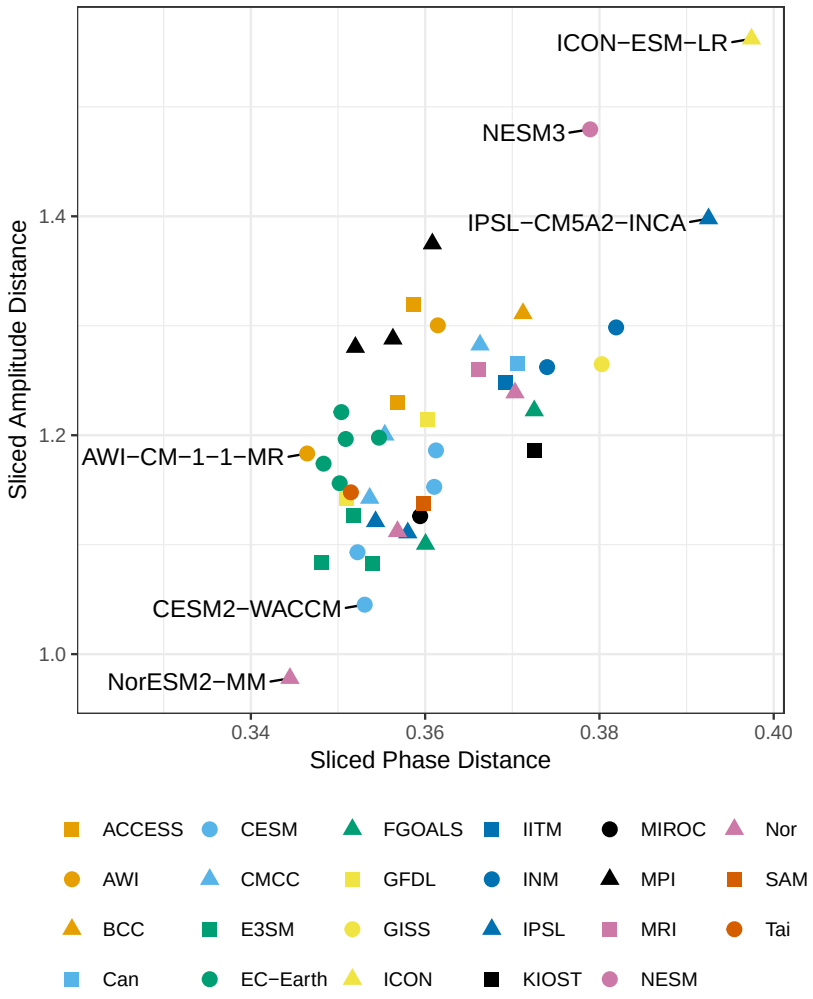
For full details on our sliced elastic distance rankings (including sliced translation distance), see Table 3.1 in Section 3.8.3. This table also includes comparisons to two common evaluation metrics for precipitation data: RMSE and MAE. Similar to sliced amplitude and sliced phase distance, RMSE and MAE rank ERA5 as most similar to GPCP. However, both RMSE and MAE rank NCEP among the average CMIP6 models. In most cases, the rankings for RMSE and MAE are similar to those of sliced amplitude distance, though there are a few notable exceptions such as the FGOALS models and ACCESS-CM2. Sliced phase distance, on the other hand, offers a mostly unique perspective on the rankings. Out of all the different rankings, only sliced phase distance ranks NCEP above all the CMIP6 models, indicating that phase variability is an important

Reanalysis	D_{SA}	D_{SP}
ERA5	0.578	0.264
NCEP	1.063	0.323

(a) Sliced amplitude and phase distances for reanalysis datasets

Percentile	D_{SA}	D_{SP}
Minimum	0.978	0.345
25th	1.138	0.352
Median	1.198	0.359
75th	1.265	0.369
Maximum	1.562	0.397

(b) Summary of sliced amplitude and phase distances for CMIP6 models



(c) Sliced amplitude and phase distances for CMIP6 Models

Figure 3.5: Sliced elastic distance from the GPCP observations to the CMIP6 model outputs and ERA5/NCEP reanalysis datasets. Table (a) shows sliced amplitude (D_{SA}) and sliced phase (D_{SP}) distances from GPCP to ERA5 and NCEP. Table (b) shows the minimum, 25th percentile, median, 75th percentile, and maximum sliced amplitude and phase distances among the CMIP6 model outputs. The scatter plot (c) has a point for each model with the x-axis and y-axis values representing the sliced phase and amplitude distances to GPCP, respectively. Models from the same group share their color and shape.

feature when determining similarity to observed data.

The sliced elastic distance provides a global assessment of a model's performance in mimicking the shape and timing of observed climatologies. However, it could be also desirable to know where and when the differences between a model and observations occur. Indeed, the intermediate results from calculating the sliced elastic distance can exactly reveal such information. We demonstrate this additional feature of our method for Indian Summer Monsoon in the following section.

3.5.2 Local Evaluation in the Indian Summer Monsoon Region

We characterize differences in phase between GPCP and the CMIP6 model outputs in terms of an influential climate component, the Indian Summer Monsoon (ISM). We focus on the Monsoon Core Region (MCR), which we define as the region of India from 15^{o}N to 30^{o}N latitude and 68^{o}E to 88^{o}E longitude. The MCR contains the areas with the highest proportion of rainfall during the monsoon season, typically assumed to be June, July, August, and September (JJAS), compared to the remaining 8 months of the year. Since the onset and retreat of summer monsoon are of fundamental importance, we focus on calculating the timing biases for these two key time points across the MCR. Various criteria have been established based on climatological precipitation or other aspects of the hydrological cycle to determine the typical onset and retreat of the monsoon season at each location [Wang et al., 2002, Fasullo and Webster, 2003, Misra et al., 2018]. We adopt a similar definition of the onset and retreat using thresholds of the maximum climatological precipitation at each location in the GPCP dataset. The onset date is calculated as the first day in the climatology that exceeds 50% of the maximum rainfall and the retreat date is the last day in the climatology that exceeds 50% of the maximum rainfall.

Figure 3.6 shows these dates for each slice location in the MCR. These maps are not intended to perfectly reproduce previous results, but rather provide us with a per-slice definition of the onset and retreat that will be used in our phase analysis. Compared to the onset and retreat dates in previous studies [e.g., Misra et al., 2018], our results are overall very similar, but exhibit more spatial smoothness due to the kernel convolution in our approach and the spatially coarser precipitation observations in our data.

For a given location in the MCR, we can take the onset and retreat dates from the maps in Figure 3.6 and then compute the timing bias, defined in (3.6), from the CMIP6 models to GPCP on those dates. Figure 3.7 demonstrates the calculation procedure for IPSL-CM5A2-INCA and ICON-ESM-LR, the two models with the highest global sliced phase distance. This demonstration focuses on a single location in the middle of the MCR (22.5^{o}N, 78^{o}E). For IPSL-CM5A2-INCA, Sepulchre et al. [2020] previously established a one-month lag in the ISM event. Their work focuses on monthly data while ours uses smoothed daily data,

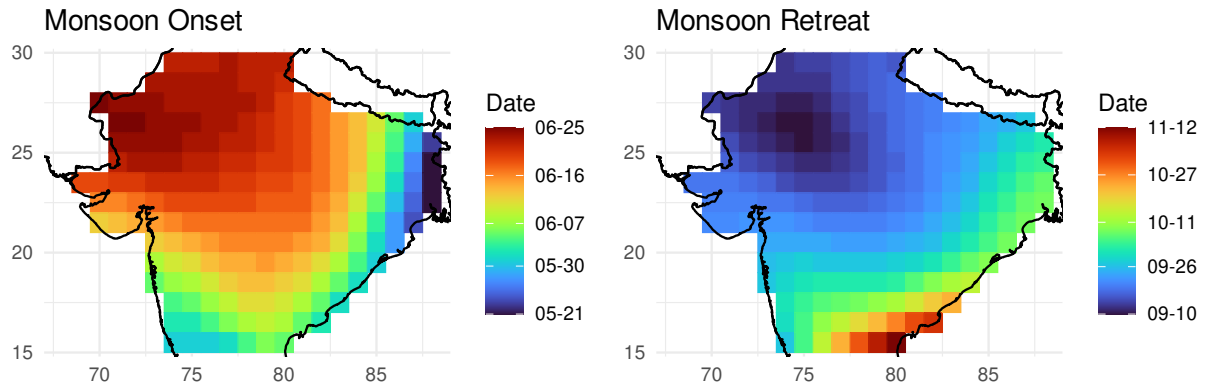


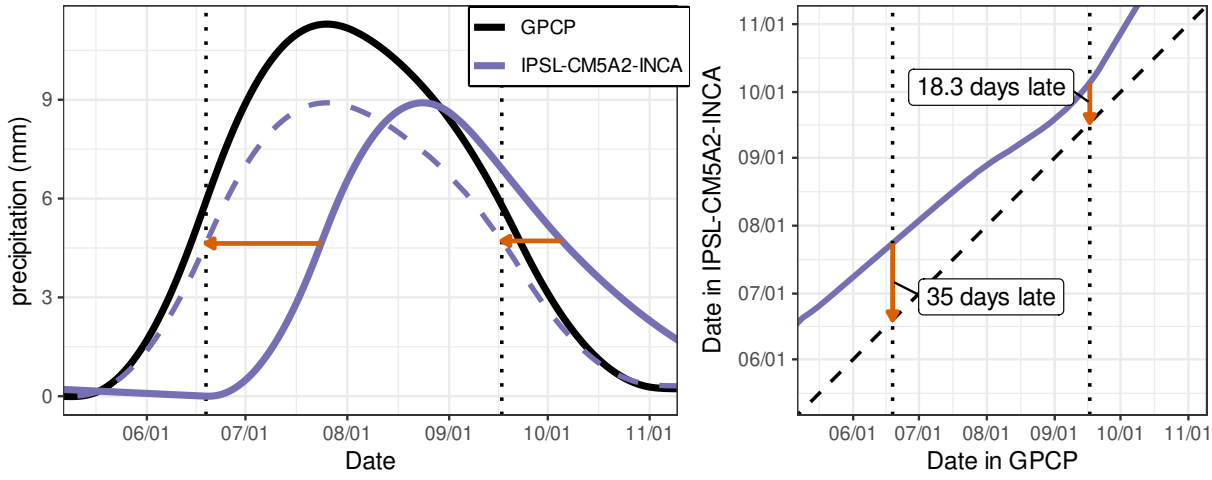
Figure 3.6: Maps of monsoon onset and retreat dates in the GPCP climatology slices. Locations within the MCR are assigned a color fill based on the local onset (left) and retreat (right) date of the Indian Summer Monsoon. Dates are determined using thresholds of 50% of the maximum precipitation in the climatology at each location.

allowing us to determine timing biases at a finer temporal resolution. At this particular location, we find that IPSL-CM5A2-INCA exhibits a late onset bias of about 35 days and a late retreat bias of about 18 days compared to GPCP, so our results generally agree with Sepulchre et al. [2020]. In contrast, ICON-ESM-LR exhibits an early onset bias of about 16 days and a late retreat bias of about 39 days.

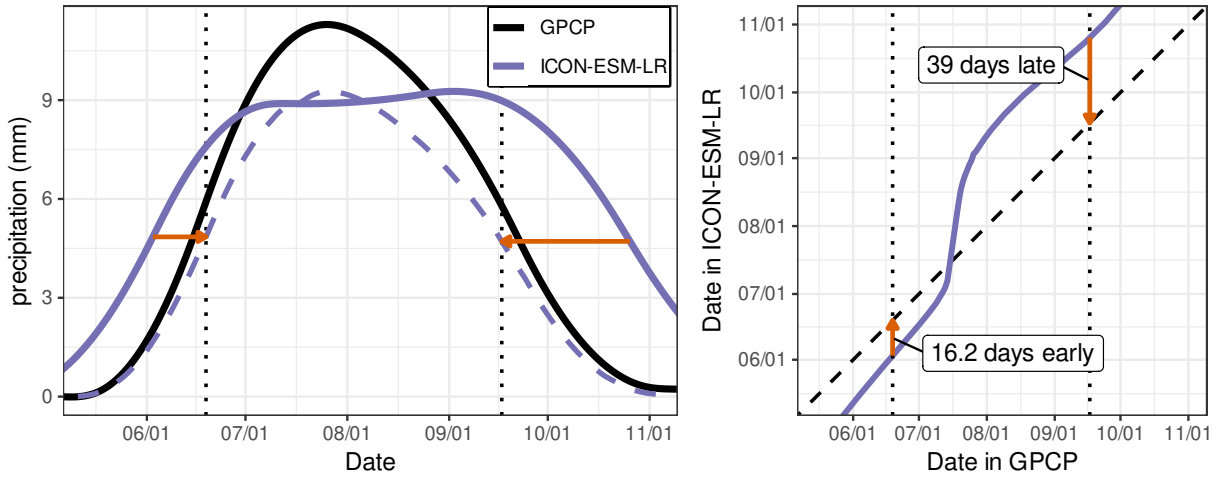
To understand spatially-varying timing biases in the onset and retreat of the Indian Summer Monsoon, we repeat the procedure in Figure 3.7 for every location in the MCR for a cohort of six CMIP6 models. The six chosen models are selected from the models with the highest and lowest sliced elastic distances in Figure 3.5. For the maps of onset timing bias, we see that NorESM2-MM, AWI-CM-1-1-MR, and CESM2-WACCM exhibit little to no timing variability in the MCR. Among these models, AWI-CM-1-1-MR seems to exhibit timing biases closest to 0 in all areas of the MCR. On the other hand, ICON-ESM-LR, IPSL-CM5A2-INCA, and NESM3 all exhibit strong biases in some or all areas of the MCR. In particular, IPSL-CM5A2-INCA shows a strong positive timing bias of 4 or more weeks in all regions.

For the maps of retreat timing bias, NorESM2-MM, AWI-CM-1-1-MR, and CESM2-WACCM all exhibit greater levels of timing bias compared to the onset event. For NorESM2-MM and AWI-CM-1-1-MR, there is a mix of small timing biases both in terms of early and late retreat, while CESM2-WACCM features mostly early or neutral timing biases. The remaining models exhibit mainly late retreat biases. For ICON-ESM-LR, late retreat biases occur in almost all areas of the MCR, while for IPSL-CM5A2-INCA and NESM3, late retreat biases occur mainly towards the western edge and center of the MCR with little to no bias elsewhere.

Overall, our results clearly characterize timing biases in the onset and retreat of the Indian Summer Monsoon, providing climate modelers with a useful diagnostic tool to understand a model's performance in

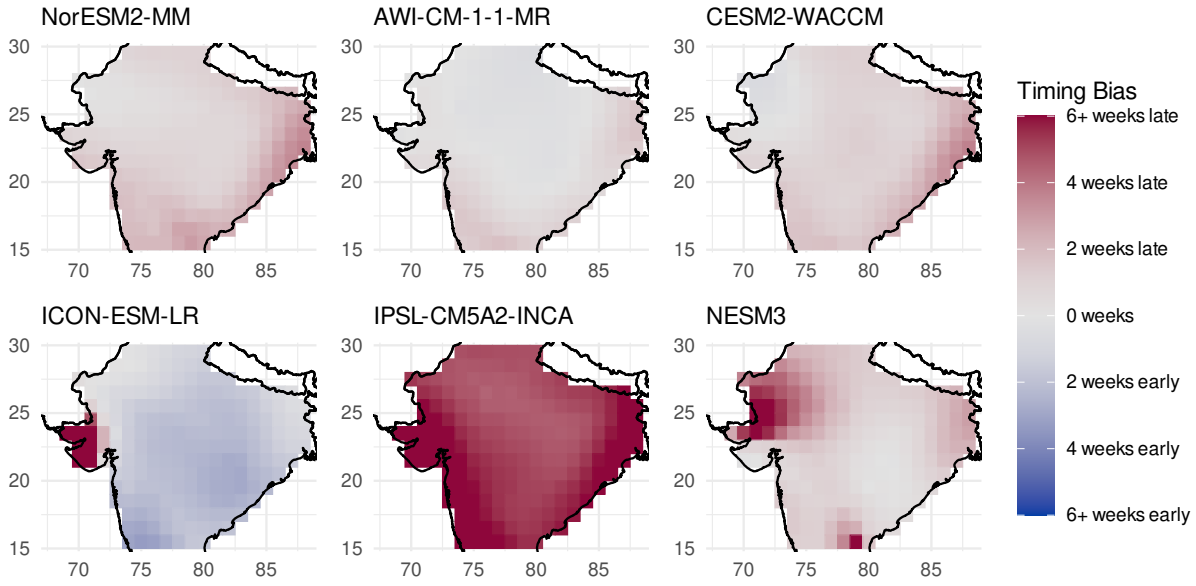


(a) GPCP and IPSL-CM5A2-INCA Slice Comparison

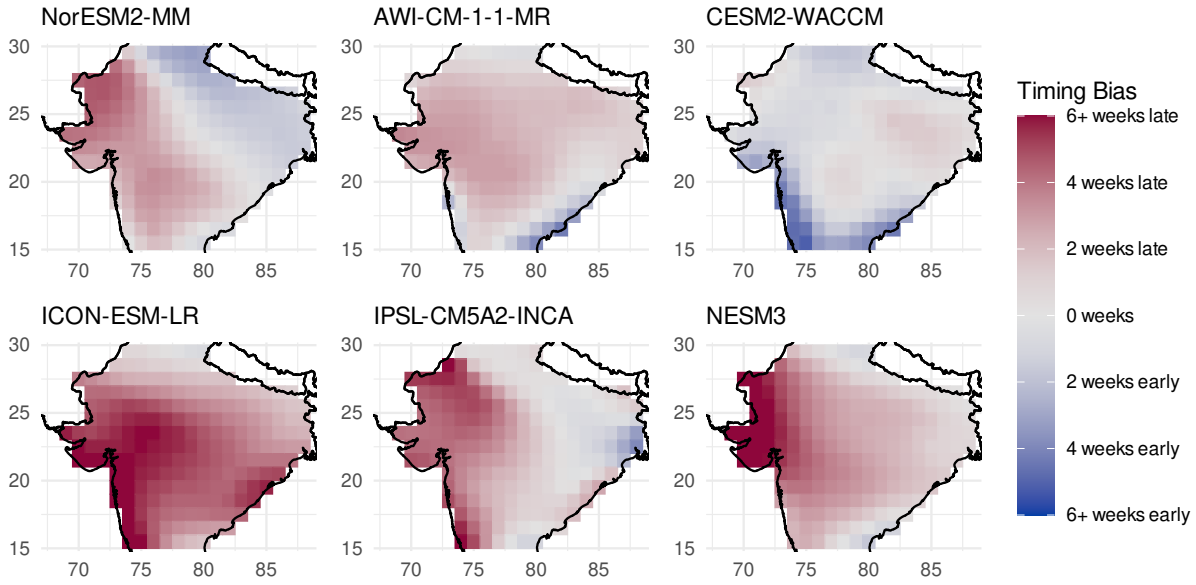


(b) GPCP and ICON-ESM-LR Slice Comparison

Figure 3.7: Timing biases from IPSL-CM5A2-INCA and ICON-ESM-LR to GPCP for the ISM onset and retreat dates. In both (a) and (b), the left plot shows the climate model and GPCP slices located at (22.5°N, 78°E). Solid lines represent the slices, with the model in purple and GPCP in black. Dotted vertical lines represent the ISM onset (June 19th) and retreat (September 17th) dates at the selected location in the GPCP data. The purple dashed lines represent the model slices after being aligned to GPCP. The plots on the right show the warping function used for alignment. The orange arrows represent the time warping applied to the model at the ISM onset and retreat dates. The direction of the arrow indicates warping forward/backward in time, and the arrow's length represents the number of days.



(a) Maps of onset timing bias



(b) Maps of retreat timing bias

Figure 3.8: Maps of timing bias for the ISM onset and retreat dates from six CMIP6 models compared to GPCP. At each location, the onset and retreat dates are determined for GPCP using the maps in Figure 3.6. Given those dates, timing biases are calculated at each slice location using the process detailed in Figure 3.7. The color fill represents the timing bias for each slice, with panel (a) containing the onset biases and panel (b) containing the retreat biases.

phase variability. Such information is not available when one simply uses summary statistics such as the mean JJAS rainfall to evaluate the skill of climate model in simulating monsoon. Furthermore, the timing biases shown in Figure 3.8, particularly the late retreat biases, indicate that some models simulate a significant amount of monsoon precipitation outside of the JJAS time interval. So, a model with a low JJAS mean may not actually underestimate the ISM precipitation, but instead mistime the event.

3.6 Discussion

To aid climate model evaluation efforts for precipitation, we developed a new metric called sliced elastic distance for quantifying model performance in a way that accounts for both spatial and temporal variability. The sliced elastic distance is a vector-valued metric that represents the amplitude, phase, and translation distance between two spatiotemporal processes. It extends the ordinary elastic distance by allowing for spatially varying time warping via a kernel convolution procedure. We focus on amplitude and phase in evaluating precipitation models. Amplitude variability corresponds to differences in the underlying shape of precipitation trajectories, while phase variability corresponds to errors in the timing of seasonal transitions and weather events such as monsoons. Distinguishing these two components provides us with more insight into the sources of model misfit. By applying our method to evaluate CMIP6 precipitation models, we quantified how each model performs using their amplitude and phase distance to observed data. We also examined the performance of CMIP6 models in capturing the proper onset and retreat timing of the Indian summer monsoon at a local spatial scale.

One limitation of our method is that the slicing process does not include a formal alignment step for the spatial domain as there is for the time domain. Spatial warping was used in Levy et al. [2014] to correct errors in the location of precipitation events in climate models, but adapting this approach for our daily analysis would be computationally infeasible. Another limitation is that the method assumes that the time domain is the unit interval $[0, 1]$, representing the climatologies for January 1 through December 31. Because climatologies are computed as long-term means over many years, a cyclical temporal domain, such as the unit sphere $\mathbb{S}^1$, may better represent the temporal structure. While such closed temporal domains are commonly used in elastic FDA [Joshi et al., 2007], only the properties of the amplitude (shape) distance have been established on this domain [Srivastava and Klassen, 2016]. Lastly, although our threshold-based approach to determining local onset and retreat dates is convenient for our shape analysis-driven approach, we understand that there are alternative approaches for defining local onset and retreat dates of monsoon seasons which may suggest different dates [i.e., Misra et al., 2018]. Timing biases based on those dates could produce slightly different results.

Our method was developed in the context of evaluating precipitation models. However, our method can be applied to compare any two spatiotemporal fields. For example, the sliced elastic distance can also be used to evaluate climate models for temperatures, or for climate model tuning [Hourdin et al., 2017]. Outside of climate science, possible applications include comparing spectral data from remote sensing data products, analyzing the functional measurements obtained from atomic-scale microscopy in material science, or video processing, where the considerations made for spherical data in our method would apply directly to 360^o videos.

3.7 Acknowledgements

We would like to thank Robert Krueger for the interesting discussions on convolution theorems and spherical harmonics. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. NCEP/DOE Reanalysis II data provided by the NOAA PSL, Boulder, Colorado, USA, from their website at `https://psl.noaa.gov`. This work is partially supported by the National Science Foundation grants NSF-DMS-1830312, NSF-DGE-1922758, and NSF-DMS-2124576 as well as the National Oceanic and Atmospheric Administration (NOAA) grant NA18OAR4310271.

3.8 Supplementary Material

3.8.1 Proof of Theorem 3.1

First we prove three useful Lemmas, including Lemma 3.1 which was previously stated in Section 3.3.3.

Lemma 3.1. *$D_E(f, g)$ is a vector-valued metric on $\mathcal{F}_\mathcal{T}$.*

Proof. Proposition 2.1 in [Jachymski and Klima, 2016] provides a clear structure for our proof. First, we show that amplitude, phase, and translation distance are each a pseudometric on $\mathcal{F}_\mathcal{T}$. Then, we finish the proof by showing the final property: for any $f, g \in \mathcal{F}_\mathcal{T}$, $f \neq g \Rightarrow D_A(f, g) > 0$, $D_P(f, g) > 0$, or $D_T(f, g) > 0$.

Section 4.10.1 in Srivastava and Klassen [2016] shows that amplitude distance is a proper metric on the quotient space $\mathbb{L}_2/\Gamma$. Since for all functions $f, g \in \mathcal{F}_\mathcal{T}$, there exist unique orbits $[q_f], [q_g] \in \mathbb{L}_2/\Gamma$, the Identity, Symmetry, and Triangle Inequality properties hold trivially on $\mathcal{F}_\mathcal{T}$ as well.

Section 4.10.2 in [Srivastava and Klassen, 2016] shows that for two functions $f, g \in \mathcal{F}_{\mathcal{T}}$, phase distance satisfies the Identity, Symmetry, and Triangle Inequality properties on $\mathcal{F}_{\mathcal{T}}$. So, phase distance is a pseudometric on $\mathcal{F}_{\mathcal{T}}$.

For translation distance, let $f, g, h \in \mathcal{F}_{\mathcal{T}}$. Note $f(0), g(0), h(0) \in \mathbb{R}$. Then,

1. $D_T(f, f) = |f(0) - f(0)| = 0$
2. $D_T(f, g) = |f(0) - g(0)| = |g(0) - f(0)| = D_T(g, f)$
3. $D_T(f, h) = |f(0) - h(0)| \leq |f(0) - g(0)| + |g(0) - h(0))| = D_T(f, g) + D_T(g, h)$.

Therefore, translation distance is a pseudometric on $\mathcal{F}_{\mathcal{T}}$.

It remains to show that for all $f, g \in \mathcal{F}_{\mathcal{T}}$, $f \neq g \Rightarrow D_A(f, g) > 0$, $D_P(f, g) > 0$, or $D_T(f, g) > 0$. We proceed by proving the contrapositive: for all $f, g \in \mathcal{F}_{\mathcal{T}}$, if $D_A(f, g) = D_P(f, g) = D_T(f, g) = 0$ then $f = g$. Let $f, g \in \mathcal{F}_{\mathcal{T}}$ with $D_A(f, g) = D_P(f, g) = D_T(f, g) = 0$. Denote the relative phase of f with respect to g as (γ_f^*, γ_g^*). $D_P(f, g) = 0 \Rightarrow \gamma_f^* = \gamma_g^*$. So,

$$
\begin{aligned}
0 &= D_A(f, g) && (3.8)\\
&= \inf_{\gamma_f, \gamma_g \in \tilde{\Gamma}_I} ||(q_f, \gamma_f) - (q_g, \gamma_g)|| && (3.9)\\
&= ||(q_f, \gamma_f^*) - (q_g, \gamma_g^*)|| && (3.10)\\
&= ||(q_f, \gamma_f^*) - (q_g, \gamma_f^*)|| && (3.11)\\
&= ||q_f - q_g||. && (3.12)
\end{aligned}
$$

The third line follows by the definition of relative phase. The last line follows by Lemma 4.2 in [Srivastava and Klassen, 2016].

It follows by property of the Euclidean norm that $q_f = q_g$. In addition, $D_T(f, g) = 0 \Rightarrow f(0) = g(0)$. So, we have $f(0) = g(0)$ and $q_f = q_g$. By the absolute continuity of f and g, it follows that $f = g$. So, for all $f, g \in \mathcal{F}_{\mathcal{T}}$, $D_A(f, g) = D_P(f, g) = D_T(f, g) = 0 \Rightarrow f = g$.

We have shown that the elastic distance $D_E(f, g)$ is a family of pseudometrics on $\mathcal{F}_{\mathcal{T}}$ that jointly satisfies the positivity property on $\mathcal{F}_{\mathcal{T}}$. So, by Proposition 2.1 in [Jachymski and Klima, 2016], it follows that D_E is a vector-valued metric on $\mathcal{F}_{\mathcal{T}}$. □

Lemma 3.2 (Slicing creates valid functions of time)**.** *If $f \in \mathcal{F}_{\mathcal{S} \times \mathcal{T}}$, then the slice $f_s(t)$ as defined in Definition 3.1 is an element of $\mathcal{F}_{\mathcal{T}}$ for all $s \in \mathcal{S}$.*

Proof. Let $f \in \mathcal{F}_{\mathcal{S} \times \mathcal{T}}$, $s^* \in \mathcal{S}$, and $r > 0$. Suppose for contradiction that $f_{s^*}(t) \notin \mathcal{F}_{\mathcal{T}}$. Then, $f_{s^*}(t)$ is not

absolutely continuous, so there exists some $\epsilon^* > 0$ such that for all $\delta > 0$, there exists a finite sequence of M sub-intervals $(a_m, b_m) \in [0,1]$, $m = 1, ..., M$, with $\sum_{m=1}^{M}(b_m - a_m) < \delta$ and $\sum_{m=1}^{M} |f_{s^*}(a_m) - f_{s^*}(b_m)| \geq \epsilon^*$.

Let $m_{max} = \arg\max_{m \in 1,...,M} |f_{s^*}(a_m) - f_{s^*}(b_m)|$. It follows that $|f_{s^*}(a_{m_{max}}) - f_{s^*}(b_{m_{max}})| \geq \epsilon^*/M$. For clarity, denote $t_1 = a_{m_{max}}$ and $t_2 = b_{m_{max}}$. Note that $t_2 - t_1 < \delta$. We have:

$$\epsilon^*/M \leq |f_{s^*}(t_1) - f_{s^*}(t_2)| \tag{3.13}$$

$$= \left| \int_{\mathcal{S}} f(s, t_1) k^r_{s^*}(s) ds - \int_{\mathcal{S}} f(s, t_2) k^r_{s^*}(s) ds \right| \tag{3.14}$$

$$= \left| \int_{\mathcal{S}} \{f(s, t_1) - f(s, t_2)\} k^r_{s^*}(s) ds \right| \tag{3.15}$$

$$\leq \left[\int_{\mathcal{S}} \{f(s, t_1) - f(s, t_2)\}^2 ds \right]^{1/2} \left(\int_{\mathcal{S}} k^r_{s^*}(s)^2 ds \right)^{1/2}, \tag{3.16}$$

where the last line follows by the Cauchy-Schwarz Inequality. This implies:

$$\frac{(\epsilon^*/M)^2}{\int_{\mathcal{S}} k^r_{s^*}(s)^2 ds} \leq \int_{\mathcal{S}} \{f(s, t_1) - f(s, t_2)\}^2 ds \tag{3.17}$$

$$\leq \max_{s \in \mathcal{S}} \{f(s, t_1) - f(s, t_2)\}^2. \tag{3.18}$$

Now, taking the square root of each side, we get:

$$\frac{\epsilon^*/M}{\left\{\int_{\mathcal{S}} k^r_{s^*}(s)^2 ds\right\}^{1/2}} \leq \max_{s \in \mathcal{S}} |f(s, t_1) - f(s, t_2)|. \tag{3.19}$$

Denote the quantity on the left side of the above inequality as:

$$\epsilon = \frac{\epsilon^*/M}{\left\{\int_{S^2} k^r_{s^*}(s)^2 ds\right\}^{1/2}}. \tag{3.20}$$

Since $k^r_{s^*}(s) \geq 0$ for all $s, s^* \in \mathcal{S}$, $r > 0$ implies that $\int_{\mathcal{S}} k^r_{s^*}(s)^2 ds > 0$. So, $\epsilon > 0$. Let $s' = \arg\max_{s \in \mathcal{S}} |f(s, t_2) - f(s, t_1)|$. Now, we have found $\epsilon > 0$ such that for all $\delta > 0$, there exists a sub-interval $(t_1, t_2) \subset [0,1]$ with $t_2 - t_1 < \delta$ and:

$$|f(s', t_2) - f(s', t_1)| \geq \epsilon. \tag{3.21}$$

We have found a location $s' \in \mathcal{S}$ such that $f(s', t)$ is not an absolutely continuous function of time, so we have a contradiction of the assumption that $f \in \mathcal{F}_{\mathcal{S} \times \mathcal{T}}$. Our proof by contradiction is complete, therefore $f_s(t) \in \mathcal{F}_{\mathcal{T}}$ for all $s \in \mathcal{S}$ and $r > 0$. □

Lemma 3.3 (Slicing extends pseudometrics)**.** *If D is a pseudometric on $\mathcal{F}$ then D_S, the sliced version of D, is a pseudometric on $\mathcal{F}_{\mathcal{S}\times\mathcal{T}}$.*

Proof. Let D be a pseudometric on $\mathcal{F}_{\mathcal{T}}$. Let D_S be the sliced distance function corresponding to D following the process in Theorem 3.1. We begin by proving the identity property. Let $f(s,t) \in \mathcal{F}_{\mathcal{S}\times\mathcal{T}}$. Then,

$$D_S(f(s,t), f(s,t)) = \left\{ \int_{S^2} D(f_s(t), f_s(t))^2 ds \right\}^{1/2} \tag{3.22}$$

$$= \left\{ \int_{S^2} 0^2 ds \right\}^{1/2} \tag{3.23}$$

$$= 0. \tag{3.24}$$

The second line holds by the identity property of D. Next, we show the symmetry property. Let $f(s,t), g(s,t) \in \mathcal{F}_{\mathcal{S}\times\mathcal{T}}$. Then,

$$D_S(f(s,t), g(s,t)) = \left\{ \int_{S^2} D(f_s(t), g_s(t))^2 ds \right\}^{1/2} \tag{3.25}$$

$$= \left\{ \int_{S^2} D(g_s(t), f_s(t))^2 ds \right\}^{1/2} \tag{3.26}$$

$$= D_S(g(s,t), f(s,t)). \tag{3.27}$$

The second line holds by the symmetry property of D. Finally, we show the triangle inequality property. Let $f(s,t), g(s,t), h(s,t) \in \mathcal{F}_{\mathcal{S}\times\mathcal{T}}$. Then,

$$D_S(f(s,t), h(s,t)) = \left\{ \int_{S^2} D(f_s(t), h_s(t))^2 ds \right\}^{1/2} \tag{3.28}$$

$$\leq \left[\int_{S^2} \Big\{ D(f_s(t), g_s(t)) + D(g_s(t), h_s(t)) \Big\}^2 ds \right]^{1/2} \tag{3.29}$$

$$\leq \left\{ \int_{S^2} D(f_s(t), g_s(t))^2 ds \right\}^{1/2} + \left\{ \int_{S^2} D(g_s(t), h_s(t))^2 ds \right\}^{1/2} \tag{3.30}$$

$$= D_S(f(s,t), g(s,t)) + D_S(g(s,t), h(s,t)). \tag{3.31}$$

The second line holds by the triangle inequality property of D on $\mathcal{F}_{\mathcal{T}}$. The third line holds by the Minkowski Inequality.

Finally, we have that for a pseudometric D on $\mathcal{F}_{\mathcal{T}}$, the sliced version of D, D_S, satisfies all three properties of a pseudometric on $\mathcal{F}_{\mathcal{S}\times\mathcal{T}}$. Therefore, if D is a pseudometric on $\mathcal{F}_{\mathcal{T}}$, then D_S is a pseudometric on $\mathcal{F}_{\mathcal{S}\times\mathcal{T}}$. □

Using the previous lemmas along with the convolution theorem from [Driscoll and Healy, 1994], we can

prove the previously stated Theorem 3.1.

Theorem 3.1. *If $D = [D_1, ..., D_m]^T$ is a vector-valued metric on $\mathcal{F}_\mathcal{T}$, and $f_s(t)$ and $g_s(t)$ are respectively the slice functions of $f(u,t) \in \mathcal{F}_{\mathcal{S}\times\mathcal{T}}$ and $g(u,t) \in \mathcal{F}_{\mathcal{S}\times\mathcal{T}}$ using a spatially continuous kernel $k(u;\theta)$ with an positive spectral density on $\mathcal{S}$ as defined in (3.3), then the vector-valued function $D_S = [D_{S1}, ..., D_{Sm}]^T$ with each component defined as*

$$D_{Si}(f,g) = \left\{ \int_\mathcal{S} D_i \left(f_s, g_s\right)^2 ds \right\}^{1/2}, \quad i = 1, ..., m,$$

is a vector-valued metric on $\mathcal{F}_{\mathcal{S}\times\mathcal{T}}$.

Proof. Let D be a vector-valued metric on $\mathcal{F}_\mathcal{T}$. For each $i \in \{1, ..., m\}$, let D_{Si} be the sliced version of D_i using kernel $k(u;\theta)$. By Lemma 3.3, $D_S = (D_{S1}, ..., D_{Sm})$ is a family of pseudometrics on $\mathcal{F}_{\mathcal{S}\times\mathcal{T}}$. To show D_S is a vector-valued metric on $\mathcal{F}_{\mathcal{S}\times\mathcal{T}}$, it suffices to show that for any $f, g \in \mathcal{F}_{\mathcal{S}\times\mathcal{T}}$, if $f \neq g$ then $D_{Si}(f,g) > 0$ for some $i \in \{1, ..., m\}$. We proceed by proving the contrapositive: for all $f, g \in \mathcal{F}_{\mathcal{S}\times\mathcal{T}}$, if $D_S(f,g) = 0_m$ then $f = g$.

Let $f, g \in \mathcal{F}_{\mathcal{S}\times\mathcal{T}}$ with $D_S(f,g) = 0_m$. Then $0 = \int_\mathcal{S} D_{Si}(f_s, g_s)ds$ for all $i \in \{1, ..., n\}$. So, for all $i \in \{1, ..., n\}$, $D_{Si}(f_s, g_s) = 0$ for almost every $s \in \mathcal{S}$. By property of vector-valued metrics, this implies that $f_s(t) = g_s(t)$ for almost every $s \in \mathcal{S}$. Let $h(s,t) = f(s,t) - g(s,t)$. Note that $h_s(t) = \int_\mathcal{S}\{f(u,t) - g(u,t)\}k_s(u;\theta)du = f_s(t) - g_s(t)$. So, for all $t \in \mathcal{T}$, $h_s(t) = 0$ for almost every $s \in \mathcal{S}$.

Now, fix $t \in \mathcal{T}$ and define the spatial convolution of h with k at time t as $c_{h,t}(s) = \int_\mathcal{S} h(u,t)k_s(u;\theta)du$. This convolution is a function of space only, serving as the spatial version of the previously defined slice functions, which are functions of time only. Note that $c_{h,t}(s)$ is equal to 0 for almost every $s \in \mathcal{S}$ because $h_s(t) = 0$ for almost every $s \in \mathcal{S}$. Additionally, since $c_{h,t}(s)$ is defined as a convolution of continuous functions on $\mathcal{S}$, it is itself a continuous function on $\mathcal{S}$. It follows by property of continuity that $c_h(s,t) = 0$ for all $s \in \mathcal{S}$.

Using the spherical harmonics representation of $c_{h,t}(s)$ [Driscoll and Healy, 1994], we can represent $c_{h,t}(s)$ as

$$c_{h,t}(s) = \sum_{l\geq 0} \sum_{|m|\leq l} \widetilde{c_{h,t}}(l,m) Y_l^m(s). \tag{3.32}$$

Where $Y_l^m(s)$ are the spherical harmonics bases and $\widetilde{c_{h,t}}(l,m)$ are the spherical harmonics coefficients for $c_{h,t}(s)$. Since the bases $Y_l^m(s)$ are orthonormal and $c_{h,t}(s) = 0$, we have that $\widetilde{c_{h,t}}(l,m) = 0$ for all l and m.

Since $c_h(s,t) = \int_\mathcal{S} h(u,t)k_s(u;\theta)du$ is a convolution of functions on $\mathcal{S}$, where $\mathcal{S}$ was previously defined to be the 2-dimensional unit sphere, using Theorem 1 in [Driscoll and Healy, 1994] we can write the spherical harmonics coefficients $\widetilde{c_{h,t}}(l,m)$ in terms of the spherical harmonics coefficients for $h(u,t)$ and $k_s(u;\theta)$,

denoted respectively as $\widetilde{h^t}(l,m)$ and $\tilde{k}(l,m)$:

$$\widetilde{c_{h,t}}(l,m) = \alpha(l)\widetilde{h^t}(l,m)\tilde{k}(l,0), \tag{3.33}$$

where $\alpha(l) = 2\pi\sqrt{\frac{4\pi}{2l+1}}$. Clearly $\alpha(l) > 0$ for all $l \geq 0$. By our assumption that $k(u;\theta)$ has positive spectral density on $\mathcal{S}$, we know that $\tilde{k}(l,0) > 0$ for all $l \geq 0$. Therefore, since $\widetilde{c_{h,t}}(l,m) = 0$, we must have $\widetilde{h^t}(l,m) = 0$ for all l and m. Using the spherical harmonics representation for $h(s,t)$, we can see:

$$h(s,t) = \sum_{l\geq 0}\sum_{|m|\leq l} \widetilde{h^t}(l,m)Y_l^m(s) \tag{3.34}$$

$$= \sum_{l\geq 0}\sum_{|m|\leq l} 0 * Y_l^m(s) \tag{3.35}$$

$$= 0. \tag{3.36}$$

So, $h(s,t) = 0$ for all $s \in \mathcal{S}$. Since t was fixed arbitrarily, we also have that $h(s,t) = 0$ for all $t \in \mathcal{T}$. Therefore, $0 = h(s,t) = f(s,t) - g(s,t)$, implying that $f = g$.

We have shown that for all $f, g \in \mathcal{F}_{\mathcal{S}\times\mathcal{T}}$, if $D_S(f,g) = 0_m$, then $f = g$. Our proof by contrapositive is complete, therefore for all $f, g \in \mathcal{F}_{\mathcal{S}\times\mathcal{T}}$, if $f \neq g$, then $D_{Si}(f,g) > 0$ for some $i \in \{1,...,n\}$. So, for any vector-valued metric D on $\mathcal{F}$, D_S is a family of pseudometrics that jointly satisfies the positivity property on $\mathcal{F}_{\mathcal{S}\times\mathcal{T}}$. Therefore, by Proposition 2.1 in [Jachymski and Klima, 2016], if D is a vector-valued metric on $\mathcal{F}_{\mathcal{T}}$ and $k(u;\theta)$ is a continuous spatial kernel with positive spectral density on $\mathcal{S}$, then D_S, the sliced version of D defined using $k(u;\theta)$, is a vector-valued metric on $\mathcal{F}_{\mathcal{S}\times\mathcal{T}}$. □

3.8.2 Sliced Elastic Distance Implementation

Before computing the sliced elastic distance, we estimate continuous functional data from the precipitation climatology at each location by applying quadratic trend filtering via the `glmgen` `R` package using a smoothing parameter of $\lambda = 1,250$ [Tibshirani, 2014]. Because climate model output and observational data are discrete in both the space and time dimensions, the exact integrals in Definition 3.1 cannot be computed. We approximate these integrals with summations and averages over a discrete set of locations. It is the user's decision to choose locations at which they want to have the slice functions. This choice may depend on the grid density of data products and preferences for the spatial resolution of slices. In our simulation and data analysis, we choose a regular latitude-longitude grid, G, resolved by 180 latitude values and 360 longitude values, to represent the spatial domain. The resolution of G is chosen to match the resolution of the GPCP data, but users can make their own choice for the grid size and structure. Given the two daily precipitation

fields represented by $f(u,t) \in \mathcal{F}_{\mathcal{S}\times\mathcal{T}}$ and $g(u,t) \in \mathcal{F}_{\mathcal{S}\times\mathcal{T}}$ and the common grid G, the sliced elastic distance between f and g can be easily obtained through the following major steps:

1. **Slices** Compute the slice functions f_s and g_s for each location $s \in G$ through multiplication of $f(u,t)$ and $g(u,t)$ with the kernel $k_s(u)$ at each time point.

2. **Elastic distance** At each location $s \in G$, compute approximate amplitude and phase distances $\tilde{D}_A(f_s, g_s)$ and $\tilde{D}_P(f_s, g_s)$ using the dynamic programming algorithm provided in R package `fdasrvf` [Tucker, 2020].

3. **Spatial weights** Assign weights for each location $s \in G$, denoted as w_s, as the cosine of its latitude to adjust for the different areas of each grid cell. This follows the standard practice for global climate data, e.g., Li et al. [2016].

4. **Sliced elastic distance** Compute the approximate sliced elastic distance $\tilde{D}_{SE}$ between f and g as follows:

$$\tilde{D}_{SE}(f,g) \approx \begin{bmatrix} \left\{W^{-1}\sum_{s\in G} w_s \tilde{D}_A(f_s,g_s)^2\right\}^{1/2} \\ \left\{W^{-1}\sum_{s\in G} w_s \tilde{D}_P(f_s,g_s)^2\right\}^{1/2} \\ \left\{W^{-1}\sum_{s\in G} w_s D_T(f_s,g_s)^2\right\}^{1/2} \end{bmatrix}, \tag{3.37}$$

   where $W^{-1} = 1/\sum_{s\in G} w_s$.

See the supplemental material for our full code implementation in `R`. To produce the spatial maps, region means, and timing biases discussed in Section 3.3, intermediate values for $\tilde{D}_A(f_s,g_s), \tilde{D}_P(f_s,g_s), D_T(f_s,g_s)$ and the relative phase functions $\gamma^*_{f_s}, \gamma^*_{g_s}$ are saved at each location $s \in G$. Note that when calculating the elastic distance between two functions $f, g \in \mathcal{F}_{\mathcal{T}}$, the `fdasrvf` implementation assumes that $\gamma^*_f(t) = I(t) = t$, the identity warping function, for identifiability.

3.8.3 Table of Global Rankings

	Sliced Amplitude	Sliced Phase	Sliced Translation	RMSE	MAE
ERA5	0.578	0.264	0.759	1.467	1.065
NorESM2-MM	0.978	0.344	0.940	1.945	1.443
CESM2-WACCM	1.045	0.353	1.278	2.012	1.493
NCEP	1.063	0.323	1.605	2.148	1.573
E3SM-2-0	1.083	0.354	1.350	2.094	1.543
E3SM-2-0-NARRM	1.084	0.348	1.341	2.105	1.550
CESM2	1.093	0.352	1.295	2.064	1.524
FGOALS-f3-L	1.101	0.360	1.448	2.442	1.777
IPSL-CM6A-LR	1.111	0.358	1.414	2.318	1.707
NorESM2-LM	1.112	0.357	1.330	2.067	1.544
IPSL-CM6A-LR-INCA	1.121	0.354	1.466	2.337	1.718
MIROC6	1.126	0.359	1.379	2.198	1.624
E3SM-1-0	1.126	0.352	1.421	2.146	1.605
SAM0-UNICON	1.138	0.360	1.524	2.210	1.635
GFDL-CM4	1.142	0.351	1.341	2.169	1.561
CMCC-ESM2	1.143	0.354	1.466	2.138	1.593
TaiESM1	1.148	0.351	1.437	2.123	1.577
CESM2-FV2	1.153	0.361	1.578	2.076	1.555
EC-Earth3	1.156	0.350	1.322	2.088	1.548
EC-Earth3-Veg	1.174	0.348	1.336	2.092	1.550
AWI-CM-1-1-MR	1.183	0.346	1.548	2.178	1.626
CESM2-WACCM-FV2	1.186	0.361	1.593	2.117	1.592
KIOST-ESM	1.186	0.372	1.535	2.201	1.649
EC-Earth3-AerChem	1.197	0.351	1.292	2.115	1.572
EC-Earth3-CC	1.198	0.355	1.348	2.120	1.570
CMCC-CM2-SR5	1.200	0.355	1.507	2.171	1.615
GFDL-ESM4	1.214	0.360	1.300	2.257	1.625
EC-Earth3-Veg-LR	1.221	0.350	1.362	2.086	1.546
FGOALS-g3	1.222	0.372	1.657	2.434	1.798
ACCESS-CM2	1.230	0.357	1.626	2.441	1.781
NorCPM1	1.239	0.370	1.616	2.124	1.595
IITM-ESM	1.248	0.369	1.662	2.303	1.730
MRI-ESM2-0	1.260	0.366	1.730	2.242	1.666
INM-CM5-0	1.262	0.374	1.594	2.276	1.709
GISS-E2-2-G	1.265	0.380	1.623	2.325	1.763
CanESM5	1.265	0.371	1.651	2.314	1.680
MPI-ESM1-2-HR	1.280	0.352	1.560	2.212	1.634
CMCC-CM2-HR4	1.282	0.366	1.510	2.232	1.640
MPI-ESM1-2-LR	1.288	0.356	1.544	2.170	1.618
INM-CM4-8	1.298	0.382	1.658	2.261	1.723
AWI-ESM-1-1-LR	1.300	0.361	1.822	2.228	1.682
BCC-ESM1	1.311	0.371	1.650	2.381	1.728
ACCESS-ESM1-5	1.320	0.359	1.882	2.409	1.775
MPI-ESM-1-2-HAM	1.375	0.361	1.810	2.329	1.752
IPSL-CM5A2-INCA	1.398	0.393	1.738	2.308	1.743
NESM3	1.479	0.379	1.840	2.337	1.719
ICON-ESM-LR	1.562	0.397	2.150	2.436	1.823

Table 3.1: CMIP6 daily precipitation models ranked based on similarity to GPCP. The ERA5 and NCEP Reanalyses are included with the CMIP6 models as a baseline. Distances are calculated between the climatologies using using sliced amplitude, phase, and translation distance (750km range) as well as RMSE and MAE. Color fill is used for visual comparison of the rankings from each distance, with yellow representing a low rank and red representing a high rank.

Chapter 4

A Multivariate Space-Time Dynamic Model for Characterizing the Atmospheric Impacts Following the Mt. Pinatubo Eruption

4.1 Introduction

[1]Anthropogenic climate change has caused long-term shifts in weather patterns, such as the significant warming of the Earth's surface and an increase in extreme events such as droughts and wildfires. Recently, many methods have been proposed to combat the effects of climate change, including mitigation techniques such as cutting emissions and climate intervention techniques such as solar radiation modification (SRM). A promising form of SRM is stratospheric aerosol injection (SAI), discussed in Robock et al. [2009], which proposes to inject large amounts of aerosols into the stratosphere with the goal of cooling surface temperatures by changing the Earth's radiation balance. Although no large-scale SAI deployment has been demonstrated, there is a need to anticipate and understand the impacts and risks of such an event. Because of this, researchers have studied natural events such as volcanic eruptions, which can have large impacts on atmospheric aerosol levels, as analogs for SAI [Robock, 2000, Rasch et al., 2008]. The June 1991 Mt. Pinatubo eruption, for example, resulted in massive quantities of sulfate aerosols in the stratosphere, which absorbed radiation

[1]This chapter is based on the following manuscript in preparation: Garrett, R. C., Shand, L., & Huerta, G. (2024). A Multivariate Space-Time Dynamic Model for Characterizing the Atmospheric Impacts Following the Mt. Pinatubo Eruption.

and thus warmed the stratosphere and cooled the Earth's surface nearly 2 years later [McCormick et al., 1995, Boretti, 2024]. The motivation of this work is to develop a statistical method that can characterize the multivariate and dynamic nature of the global climate impacts following the Mt. Pinatubo eruption. Understanding the spatial and temporal characteristics of interactions between atmospheric processes affected by the eruption is important for anticipating the near to far-term impacts of future events. Figure 4.1, for example, shows stark changes in the standardized anomalies for three climate variables affected by the eruption: aerosol optical depth (AOD), upwelling longwave radiation, and stratospheric temperatures. There is a known physical dependency between these variables which is especially clear in the similar spatial patterns of AOD and stratospheric temperature anomalies.

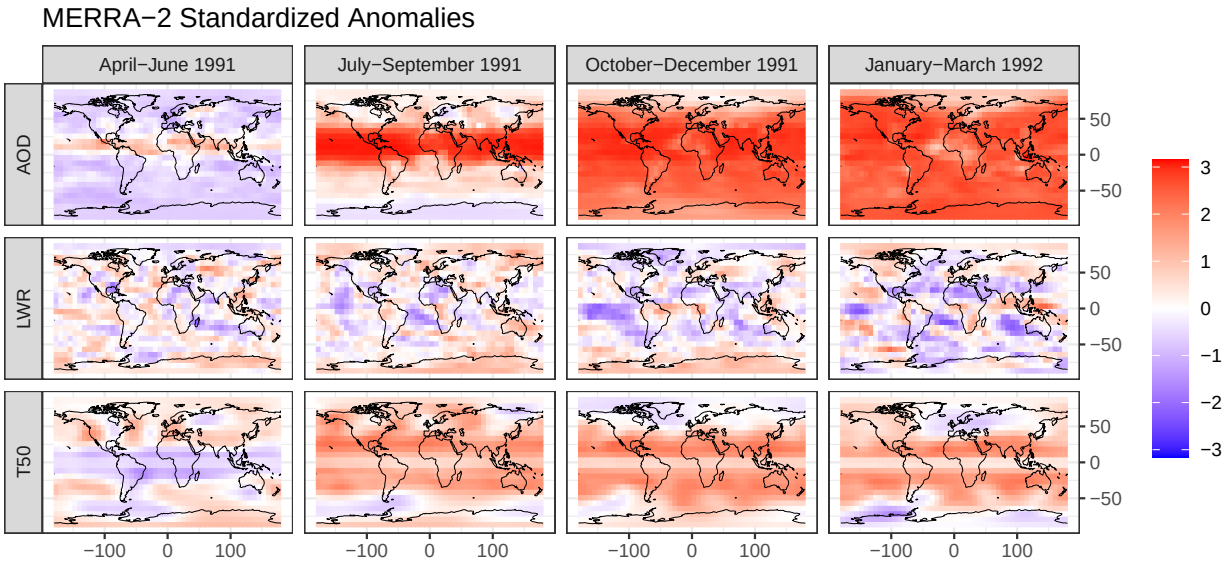


Figure 4.1: Quarterly mean summaries of aerosol optical depth (AOD), longwave radiation (LWR), and stratospheric temperature at the 50mb pressure level (T50) for one year following the Mt. Pinatubo Eruption. Each variable is obtained from the MERRA-2 Reanalysis dataset and converted to standardized anomalies as detailed in Section 4.2 before taking means over the indicated months. Color fill is used to denote the standardized anomaly at each location.

Many approaches have been developed in the spatial statistics literature that are scalable to large-scale climate data. One popular approach is Fixed Rank Kriging [e.g. Cressie and Johannesson, 2008, Sainsbury-Dale et al., 2024], which relies on a discrete approximation of the underlying continuous spatial covariance. Other fixed rank approaches include spatial basis function representations such as Lattice Kriging [Nychka et al., 2015] and Multi-Resolution Approximations [Katzfuss and Gong, 2020], which are summarized along with others in [Cressie et al., 2022]. These approaches use the basis function representation to reduce the dimension of the spatial process and induce sparsity in the covariance or precision matrix. Finally, some approaches, such as the Nearest Neighbor Gaussian Process [Datta et al., 2016, Finley et al., 2018],

utilize the Vecchia approximation [Vecchia, 1988], which induces sparsity in the covariance matrix using a nearest neighbor representation, to estimate latent Gaussian processes. More recently, this work has been generalized to enable more accurate approximations of Gaussian processes in Katzfuss and Guinness [2021]. A comprehensive review and comparison of the above methods can be found in Heaton et al. [2019].

Some of these methods have been extended to model space-time data using a state-space framework and/or Bayesian hierarchical modeling [see Banerjee et al., 2014]. These include Stroud et al. [2001], Huerta et al. [2004] and Gelfand et al. [2005], which were some of the first to consider a dynamic modeling framework for space-time processes. Finley et al. [2012] proposes a Bayesian dynamic model for space-time data using predictive processes [Finley et al., 2009]. Katzfuss and Cressie [2012] proposes a large-scale spatiotemporal smoothing method using a multiresolution basis representation. Jurek and Katzfuss [2023] presents a fast approximation to the Forward Filter, Backwards Sampling (FFBS) algorithm commonly used for estimating dynamic spatiotemporal models. See Cressie and Wikle [2011] for a thorough review of the literature on dynamic spatiotemporal models.

Fewer available methods can handle multivariate space-time processes due to the high-dimensional nature of such processes and the tendency for model specifications to become over parameterized. Gelfand et al. [2005] propose a dynamic model to capture the joint variability of temperature and precipitation over space and time. Although this approach is capable of capturing multivariate dependencies in the spatial random effect, these dependencies are not allowed to vary across space. Bradley et al. [2015] proposed a multivariate space-time model for high-dimensional areal data while Kleiber et al. [2019] proposed a multiresolution basis function approach to ensure scalability while capturing multivariate correlations in the spatial covariance matrix. Some multivariate spatial methods are very challenging to extend to spatiotemporal models due to their high computational demand, for example the cokriging methods and models of coregionalization discussed in Gelfand [2021]. Banerjee et al. [2008] proposed a Bayesian multivariate predictive process model for spatial data and suggest its scalability to space-time modeling.

The impact of stratospheric aerosols on the Earth's radiative balance and temperatures is well documented but not well quantified due to the high complexity of climate dynamics. Due to our interest in understanding the joint evolution of atmospheric processes and the limited availability of scalable multivariate space-time methods, we propose a novel multivariate space-time dynamic model (MV-STDM) to capture basic dependencies between AOD, upwelling longwave radiation, and stratospheric temperatures impacted by the Mt. Pinatubo eruption. Our MV-STDM incorporates a spatial basis function representation for computational efficiency due to the high-dimensional nature of multivariate spatiotemporal fields data. The main novelties of our work are the multivariate framework, which is motivated by the unique application to characterizing the impacts of the Mt. Pinatubo eruption, and the estimation of a spatially-varying transition matrix

that captures temporal dependence in the basis function coefficients. We estimate our model within a Bayesian hierarchical framework with a customized Markov Chain Monte Carlo (MCMC) sampling algorithm. By fitting our model to observational data over a time period covering the Mt. Pinatubo eruption, we find evidence of significant spatial trends in the evolution of each atmospheric process. We also find that our MV-STDM offers advantages over simpler univariate spatiotemporal models for predicting changes in stratospheric temperatures following the eruption. More detailed results are given in Section 4.5.

The remainder of this paper is organized as follows. Section 4.2 presents details on the atmospheric parameters of interest to motivate our multivariate model. Section 4.3 describes our model formulation and MCMC sampling procedure used for estimation. Section 4.4 demonstrates the robustness of our model under a simulated data scenario, and Section 4.5 presents the application to the motivating dataset. Lastly, Section 4.6 provides a discussion of our findings as well as possible extensions and limitations of our work.

4.2 MERRA-2 Data

To understand the multivariate evolution of atomospheric processes following the 1991 Mt. Pinatubo eruption, we study three different climate variables. These include aerosol optical depth (AOD), upwelling longwave radiation (LWR), and stratospheric temperature measured at 50mb (T50). This collection of variables represents an important interaction between atmospheric processes following the eruption, where the influx of stratospheric aerosols absorbed a greater than usual amount of upwelling longwave (infrared) radiation, leading to global stratospheric warming which reached up to 3^oC near the equator [Angell, 1997]. We obtain each variable from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2), which is a global reanalysis dataset that includes many climate variables [Gelaro et al., 2017]. For AOD, we obtain the total aerosol extinction aerosol optical thickness variable at a 550 nm wavelength, or the product labeled TOTEXTTAU. For LWR, we obtain the upwelling longwave flux at top of atmosphere variable in W/m^2, or LWTUP. Finally, for stratospheric temperature, we obtain the air temperature at a pressure level of 50 mb, labeled as T050. All three variables were obtained at a monthly temporal frequency and on a $1^o \times 1^o$ regular latitude-longitude grid.

After obtaining the MERRA-2 datasets, we apply a few preprocessing steps to prepare the data for our model. First, the spatial fields for each variable and time point are regridded to a 24×48 spatial grid through spatial averaging. Because our approach is focused on understanding multivariate relationships rather than providing fine-scale spatial predictions, this lower resolution helps to keep the dimension of our data low while still capturing long range spatial trends. Second, we subset a twelve-year time period of 1984-1995, centered around the eruption, to have a sufficient number of time steps with which to estimate our model. Lastly,

to isolate the effects of the eruption from typical spatial or seasonal trends, we convert the observations for each variable into standardized anomalies. This first requires computing the the monthly mean and monthly standard deviation at each location over the 1984-1995 time period. Standardized anomalies are then calculated for each observation and time by subtracting the monthly mean and dividing by the monthly standard deviation at that location. Figure 4.1, referenced in the introduction, shows these anomalies over four 3-month periods before and after the eruption.

4.3 Method

4.3.1 Space-Time Framework

Dynamic linear models (DLMs) are general class of models that provide a state space representation for time series as presented in Petris et al. [2009] and West and Harrison [1997]. Adaptations of DLMs for spatiotemporal data can be found in Finley et al. [2009], Finley et al. [2012], Banerjee et al. [2014], Datta et al. [2016], Wikle et al. [2019], and others such as Tonellato [1997], Stroud et al. [2001], and Gelfand et al. [2005], Huerta et al. [2004]. Our modeling approach is motivated by Rougier et al. [2023] which discusses the Kalman Filter as a "workforce" for dynamic modeling of complex environmental processes. We begin by introducing our chosen DLM framework for spatiotemporal data, then extend this model to accommodate multivariate spatiotemporal data.

We consider the problem of modeling spatiotemporal data observed at N locations over T time periods. We denote the observations from the spatial field at time t as the $N \times 1$ vector $\boldsymbol{y}_t$. Following typical DLM methodology, we begin by specifying an *observation equation* which models each $\boldsymbol{y}_t$ using a spatial basis function representation. Our basis matrix $\boldsymbol{\Phi}$ is formed with elements from K radial basis functions (RBFs), each of which is centered on a grid covering the spatial domain. The coefficients associated with the RBFs vary at each time step t and their dependence is captured in the *evolution equation*, which captures spatial and temporal autocorrelation through the use of a transition matrix and spatially-correlated error structure. Overall, our DLM equations are written as

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \overset{\text{ind.}}{\sim} N(\mathbf{0}_N, \sigma_t^2 \boldsymbol{I}_N), \\ \boldsymbol{\alpha}_t &= A\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}_K, \tau^2(\boldsymbol{B}'\boldsymbol{B})^{-1}), \end{aligned} \tag{4.1}$$

where $\boldsymbol{I}_N$ is the $N \times N$ identity matrix, $\mathbf{0}_N$ denotes the $N \times 1$ vector filled with 0, and $'$ is the matrix transpose operation. In the observation equation, $\boldsymbol{\Phi}$ is the $N \times K$ matrix containing the RBFs, $\boldsymbol{\alpha}_t$ is the $K \times 1$ state vector of time-varying coefficients, and $\boldsymbol{\epsilon}_t$ is the $N \times 1$ vector of measurement errors at each

location with variance parameterized by σ_t^2. In the evolution equation, A is the $K \times K$ transition matrix which captures the transitions between states. To create a first-order autoregressive structure for the K coefficients in the state vector, we further constrain A to be a diagonal matrix. Lastly, $\boldsymbol{\eta}_t$ is the $K \times 1$ vector of spatially-correlated innovation errors with covariance parameterized by a pre-specified $K \times K$ spatial autoregression (SAR) matrix $\boldsymbol{B}$ and a scale parameter τ^2.

Both $\boldsymbol{\Phi}$ and $\boldsymbol{B}$ follow the specification introduced in Nychka et al. [2015] for rectangular spatial geometries and later adapted for spherical spatial geometries in the 'LatticeKrig' R package [Nychka et al., 2016]. For our application, we elect to use the spherical geometry representation to cover the global spatial domain of our data. Under this specification, our basis consists of K Wendland RBFs each centered at a different location on an icosahedral grid over the Earth. We denote this grid as the 'basis grid'. Following Nychka et al. [2015], each RBF is created using the following Wendland function

$$\phi(d) = \begin{cases} (1-d)^6(35d^2 + 18d + 3)/3 & 0 \leq d \leq 1, \\ 0 & d > 1, \end{cases} \tag{4.2}$$

where the distances d are computed from each of the N locations of the observations to the K locations in the basis grid. Great circle distance is used to accommodate the spherical structure, and a scaling factor is applied so that the range of each RBF extends approximately 2.5 times the distance between neighboring locations in the basis grid. For interpretability of our model parameters, particularly A, we consider only a single resolution of RBFs. Because the RBFs are associated with different locations, we capture spatial dependence within the innovation errors $\boldsymbol{\eta}_t$ by using the SAR matrix, $\boldsymbol{B}$. For spherical geometries, $\boldsymbol{B}$ is a sparse matrix with each element varying depending on the adjacency structure of the basis grid. The (i, j) element of $\boldsymbol{B}$ is equal to $1 + \kappa^2$ if $i = j$, $-1/n_i$ if the jth location in the basis grid is adjacent to the ith location, and 0 otherwise, where n_i denotes the number of locations adjacent to the ith location. Because the basis grid is created by recursively dividing the faces of an icosahedron, n_i is either 5 or 6. The κ parameter must be greater than 0 and its value determines the overall strength of the spatial correlation, with smaller values enforcing greater correlation between neighboring RBFs. See Nychka et al. [2016] for details and code implementation.

4.3.2 Multivariate Extension

We now extend our model from (4.1) to the multivariate case. For variable $i = 1, \ldots, M$, let $\boldsymbol{y}_t^{(i)}$ denote the $N \times 1$ vector of observations from the spatial field at time t. We jointly represent the observations $\boldsymbol{y}_t^{(i)}$, along with the corresponding measurement errors $\boldsymbol{\epsilon}_t^{(i)}$, state vector $\boldsymbol{\alpha}_t^{(i)}$, and innovation errors $\boldsymbol{\eta}_t^{(i)}$, using the

following structures:

$$\boldsymbol{Y}_t = \begin{bmatrix} \boldsymbol{y}_t^{(1)} \\ \vdots \\ \boldsymbol{y}_t^{(M)} \end{bmatrix}, \qquad \boldsymbol{E}_t = \begin{bmatrix} \boldsymbol{\epsilon}_t^{(1)} \\ \vdots \\ \boldsymbol{\epsilon}_t^{(M)} \end{bmatrix}, \qquad \boldsymbol{\alpha}_t = \begin{bmatrix} \boldsymbol{\alpha}_t^{(1)} \\ \vdots \\ \boldsymbol{\alpha}_t^{(M)} \end{bmatrix}, \qquad \boldsymbol{H}_t = \begin{bmatrix} \boldsymbol{\eta}_t^{(1)} \\ \vdots \\ \boldsymbol{\eta}_t^{(M)} \end{bmatrix}. \tag{4.3}$$

As a result of stacking the parameters for each variable, the observations and measurement errors, $\boldsymbol{Y}_t$ and $\boldsymbol{E}_t$, are now of size $MN \times 1$, and the state vector and innovation errors, $\boldsymbol{\alpha}_t$ and $\boldsymbol{H}_t$, are now of size $MK \times 1$. Additionally, we adjust the remaining model parameters to accommodate the increase in dimension size. The basis, measurement variance, and innovation variance matrices are expanded to match the increased model dimension as follows

$$\boldsymbol{\Phi}_M = \boldsymbol{I}_M \otimes \boldsymbol{\Phi}, \qquad \boldsymbol{V}_t = \text{diag}\left(\sigma_{1t}^2, \ldots, \sigma_{Mt}^2\right) \otimes \boldsymbol{I}_n, \qquad \boldsymbol{Q} = \text{diag}\left(\tau_1^2, \ldots, \tau_M^2\right) \otimes (\boldsymbol{B}'\boldsymbol{B})^{-1}, \tag{4.4}$$

where $\otimes$ denotes the Kronecker product and $\text{diag}\left(\sigma_{1t}^2, \ldots, \sigma_{Mt}^2\right)$, $\text{diag}\left(\tau_1^2, \ldots, \tau_M^2\right)$ denote the $M \times M$ diagonal matrices with the elements $\sigma_{1t}^2, \ldots, \sigma_{Mt}^2$ and $\tau_1^2, \ldots, \tau_M^2$, respectively. We specify the same set of radial basis functions for each variable. Thus, $\boldsymbol{\Phi}_M$ is an $MN \times MK$ basis matrix which contains a copy of the original $\boldsymbol{\Phi}$ for each variable, $\boldsymbol{V}_t$ is the $MN \times MN$ diagonal measurement variance matrix which incorporates a different scale parameter σ_{it}^2 for each variable, and $\boldsymbol{Q}$ is the $MK \times MK$ innovation covariance matrix which keeps the same spatial autoregression matrix $\boldsymbol{B}$ for each variable, but allows the scale parameter τ_i^2 to vary between variables.

Note that neither $\boldsymbol{V}_t$ or $\boldsymbol{Q}$ incorporate multivariate dependencies. This is a purposeful choice to isolate multivariate relationships in the transition matrix, $\boldsymbol{A}$, thereby creating a more interpretable model. To match the enlarged dimension of the state vector, $\boldsymbol{A}$ is now an $MK \times MK$ matrix. We partition $\boldsymbol{A}$ into blocks to separate relationships between variables as follows

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & \ldots & A_{1M} \\ \vdots & \ddots & \vdots \\ A_{M1} & \ldots & A_{MM} \end{bmatrix}, \tag{4.5}$$

where each block A_{ij} is a $K \times K$ matrix that models the impact of $\boldsymbol{\alpha}_{t-1}^{(j)}$ on $\boldsymbol{\alpha}_t^{(i)}$. Since each element of $\boldsymbol{\alpha}_t^{(i)}$ corresponds to one RBF and thus one location in the basis grid, a fully dense specification of A_{ij} would capture the dependence of variable i on variable j for every pair of locations in the basis grid. For a dense specification of the A_{ij} matrices, the resulting number of parameters in $\boldsymbol{A}$ would be $(MK)^2$, a potentially huge number if

even a modest resolution is chosen for $\boldsymbol{\Phi}$, thus presenting challenges for estimation and interpretation. To maintain sparsity in our model specification while keeping important temporal dependencies in place, we constrain each block A_{ij} for $i, j = 1, \ldots, M$ as follows

$$A_{ij} = \text{diag}(\tilde{A}_{ij}), \quad i, j \in \{1, \ldots, M\}, \tag{4.6}$$

where each $\tilde{A}_{ij}$ is a $K \times 1$ vector. This constraint reduces the overall dimension of $\boldsymbol{A}$ to M^2K, and the dependence of variable i on variable j is now captured only between state vector coefficients which correspond to the same location in the basis grid.

Finally, our multivariate space-time dynamic model (MV-STDM) incorporates the above parameterizations to provide a similar set of DLM equations to (4.1) and is written as

$$\begin{aligned} \boldsymbol{Y}_t &= \boldsymbol{\Phi}_M \boldsymbol{\alpha}_t + \boldsymbol{E}_t, \quad \boldsymbol{E}_t \overset{\text{ind.}}{\sim} N(\boldsymbol{0}_{NM}, \boldsymbol{V}_t), \\ \boldsymbol{\alpha}_t &= \boldsymbol{A}\boldsymbol{\alpha}_{t-1} + \boldsymbol{H}_t, \quad \boldsymbol{H}_t \overset{\text{i.i.d.}}{\sim} N(\boldsymbol{0}_{KM}, \boldsymbol{Q}), \end{aligned} \tag{4.7}$$

where $\boldsymbol{Y}_t$, $\boldsymbol{\alpha}_t$, $\boldsymbol{E}_t$, and $\boldsymbol{H}_t$ follow the structure outlined in (4.3), $\boldsymbol{\Phi}_M$, $\boldsymbol{V}_t$, and $\boldsymbol{Q}$ follow the specification in (4.4), and $\boldsymbol{A}$ is subject to the sparsity constraint detailed in (4.6). The MV-STDM incorporates a first-order vector autoregressive, or VAR(1), structure to capture relationships in the evolution of the state vector for each variable. Following typical assumptions for VAR models Lütkepohl [2005], both the coefficients, $\boldsymbol{A}$, and the covariance, $\boldsymbol{Q}$, of the innovation process are fixed in time. The prosposed model structure allows us to partition the total variability in the data into random sources, such as measurement errors and spatial innovations, and deterministic sources, such as multivariate and temporal dependencies in the transition matrix.

Interpretation of the blocks A_{ij} of the transition matrix $\boldsymbol{A}$ is of particular interest to our application, allowing us to assess the strength and direction of multivariate temporal dependencies. We can project A_{ij} onto the spatial domain of the observations via the product $\boldsymbol{\Phi}A_{ij}$. Due to our sparsity constraint, each A_{ij} block has only K terms corresponding to each RBF, so this allows us to visualize the relationship between $\boldsymbol{\alpha}_t^{(i)}$ and $\boldsymbol{\alpha}_{t-1}^{(j)}$ as a spatial map. To preserve the original scale of the coefficients in A_{ij}, we divide each element of $\boldsymbol{\Phi}A_{ij}$ by the corresponding element of $\boldsymbol{\Phi}\boldsymbol{I}_K$. This process is demonstrated in Figure 4.2, which depicts the blocks A_{ij} both as points on the basis grid as well as projected onto the spatial domain. In Figures 4.3, 4.4, and 4.5, we show only the $\boldsymbol{\Phi}A_{ij}$ representation as it provides a more clear depiction of the spatial trends.

4.3.3 MCMC Sampling

Estimation for the MV-STDM in (4.7) is performed using a Markov Chain Monte Carlo (MCMC) sampler. We construct a Gibbs sampler to estimate the parameters which requires deriving full conditional distributions for each model parameter in (4.7). In this section, we specify our prior distribution selections and respective posterior distributions. For ease of sampling, we specify semi-conjugate priors.

Our MV-STDM takes the standard form of a DLM, so conditional posterior sampling for the state vectors $\boldsymbol{\alpha}_0, \ldots, \boldsymbol{\alpha}_T$ and variance terms σ_{it}^2 and τ_i^2 can be performed using similar methodology to Petris et al. [2009], which relies on the Forward Filtering, Backward Sampling (FFBS) algorithm from Carter and Kohn [1994] and Frühwirth-Schnatter [1994] to sample the state vector. However, the transition matrix $\boldsymbol{A}$ has an additional sparsity constraint that is not common in the DLM literature. To account for this, we reformulate the evolution equation in (4.7) to sample the full conditional distribution of $\boldsymbol{A}$.

Following Bayesian VAR methods [Chan, 2020], but making modifications for our specific model structure, we cast the transition equation as a Bayesian linear regression problem by reorganizing the structure of some parameters. First, we stack the $\boldsymbol{\alpha}_t$ and $\boldsymbol{H}_t$ vectors for every time point to obtain response and error vectors of size $TMK \times 1$. Then, we represent the nonzero terms of $\boldsymbol{A}$ as $\tilde{\boldsymbol{A}} = (\tilde{A}_{ij})$, the $MK \times M$ block matrix with blocks $\tilde{A}_{ij}$ for $i, j \in \{1, \ldots, M\}$. We take the column-wise vectorization of $\tilde{\boldsymbol{A}}$, denoted as $\text{vec}(\tilde{\boldsymbol{A}})$, to obtain a $M^2K \times 1$ vector of regression coefficients. Lastly, we build the design matrix of the linear regression model by specifying $\boldsymbol{X}_t^{\alpha} = \boldsymbol{I}_M \otimes \begin{bmatrix} \text{diag}(\boldsymbol{\alpha}_t^{(1)}) & \ldots & \text{diag}(\boldsymbol{\alpha}_t^{(M)}) \end{bmatrix}$ to reincorporate the diagonal structure of the blocks of $\boldsymbol{A}$. The kroncker product serves to increase the dimension of $\boldsymbol{X}_t^{\alpha}$ to $MK \times M^2K$ so that the number of columns matches the size of $\text{vec}(\tilde{\boldsymbol{A}})$. We can stack $\boldsymbol{X}_0^{\alpha}, \ldots, \boldsymbol{X}_{T-1}^{\alpha}$ so that the number of rows matches the size of the stacked $\boldsymbol{\alpha}_t; t = 1, \ldots T$. Finally, the following linear regression model

$$\begin{bmatrix} \boldsymbol{\alpha}_T \\ \vdots \\ \boldsymbol{\alpha}_1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_{T-1}^{\alpha} \\ \vdots \\ \boldsymbol{X}_0^{\alpha} \end{bmatrix} \text{vec}\left(\tilde{\boldsymbol{A}}\right) + \begin{bmatrix} \boldsymbol{H}_T \\ \vdots \\ \boldsymbol{H}_1 \end{bmatrix}, \quad \begin{bmatrix} \boldsymbol{H}_T \\ \vdots \\ \boldsymbol{H}_1 \end{bmatrix} \sim N\left(\boldsymbol{0}_{TMK}, \boldsymbol{I}_T \otimes \boldsymbol{Q}\right) \tag{4.8}$$

is equivalent to the evolution equation in (4.7). We then can specify normal distributions for $\boldsymbol{\alpha}_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0)$ and the nonzero elements of $\boldsymbol{A}$, i.e.

$$\tilde{A}_{ij} \overset{ind}{\sim} \begin{cases} N(\boldsymbol{1}_K, \lambda \boldsymbol{I}_k), & i = j, \\ N(\boldsymbol{0}_K, \lambda \boldsymbol{I}_k), & i \neq j. \end{cases} \tag{4.9}$$

The assigned prior distribution for $\boldsymbol{\alpha}_0$ follows common practice for the state vector at time 0 in DLMs. The

prior distribution for $\tilde{A}_{ij}$ is a special case of the Minnesota prior, which is popular for Bayesian VAR models [Lütkepohl, 2005]. In the case that $i = j$, the prior mean is 1, which supports our belief that there will be strong positive autocorrelation for each variable over time. In the case that $i \neq j$, the prior mean is 0, which allows for either positive or negative cross-correlations over time. In practice, small values for the the prior variance, λ, serve to shrink the coefficients of $\tilde{A}_{ij}$ closer to their prior mean. The sampling of the full conditional posterior of $\boldsymbol{A}$ can be completed via Bayesian linear regression method as shown in Chan [2020] for the general Minnesota prior. The remaining prior distributions are specified as

$$\sigma_{it}^2 \overset{ind}{\sim} IG(a_\sigma, b_\sigma), \qquad \tau_i^2 \overset{ind}{\sim} IG(a_\tau, b_\tau), \tag{4.10}$$

where IG is the Inverse-Gamma distribution. These prior distributions for σ_{it}^2 and τ_i^2 allow for Normal-Inverse Gamma conjugacy to obtain the corresponding full conditional posterior distributions.

Technical details on our Gibbs sampling procedure are made available in Section 4.8.1. We implemented this procedure in Rcpp [Eddelbuettel and François, 2011] using the 'Armadillo' package for matrix operations [Eddelbuettel and Sanderson, 2014]. Sampling for the state vector $\boldsymbol{\alpha}_t$ is based on a modified version of the FFBS algorithm in the 'dlm' R package [Petris, 2010]. Their code is not intended for a high dimensional parameter space such as the one corresponding to our multivariate spatiotemporal model, so in our FFBS implementation modifications were made to account for the sparsity structure of the model parameters.

4.4 Simulation Study

4.4.1 Synthetic Dataset

A simulation study was performed to test the effectiveness of our sampling algorithm and identifiability of our model parameters. We generate a synthetic data from our MV-STDM in (4.7) with $N = 1,152$ locations, $T = 144$ time points, and $M = 3$ variables to mimic the dimension of the MERRA-2 reanalysis data described in Section 4.2. For the basis functions $\boldsymbol{\Phi}$, we construct $K = 42$ RBFs on an icosahedral grid. Chosen for ease of computation, this is a lower basis resolution than we use for the main analysis in Section 4.5. For the innovation covariance $\boldsymbol{Q}$, we specify $\tau_i^2 = 5$ for all variables $i = 1, 2, 3$, and generate the spatial autoregression matrix $\boldsymbol{B}$ using a value of $\kappa = 2$. To ensure the simulated data have sufficient spatial signal, we specify $\sigma_{it}^2 = 2$ for all variables $i = 1, 2, 3$ and all time points $t = 1, \ldots, T$ so that the measurement errors, driven by $\boldsymbol{V}_t$, have a smaller impact on the process than the innovations, which are driven by $\boldsymbol{Q}$.

For the transition matrix, $\boldsymbol{A}$, we specify varying temporal correlation strengths for the blocks A_{ij}, each of which determines the temporal dependence of variable i on variable j. Figure 4.2 shows the chosen parameter

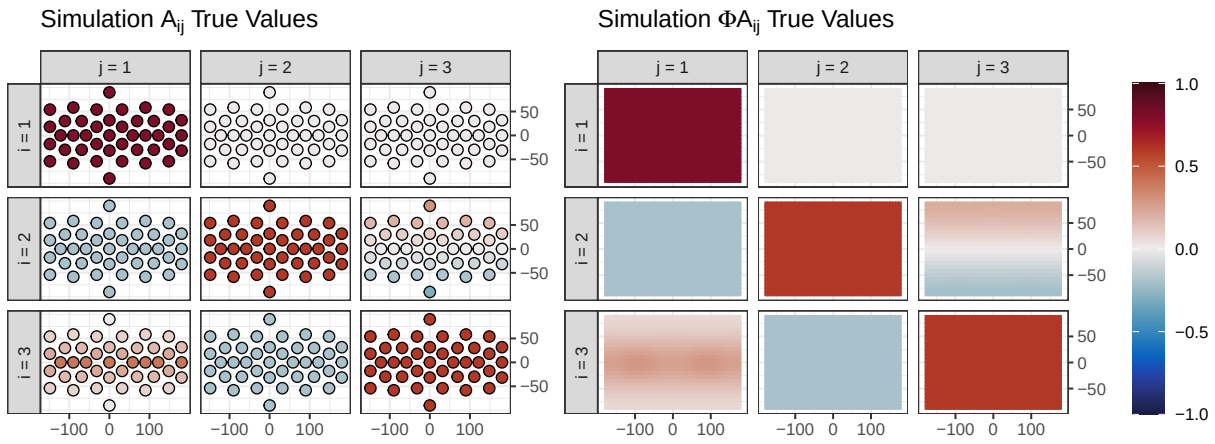


Figure 4.2: True values for the transition matrix, $\boldsymbol{A}$, used to generate our synthetic data example. The left panel shows each of the the blocks A_{ij} of $\boldsymbol{A}$, which represent the temporal dependency of variable i on variable j for each combination $i, j = 1, 2, 3$. These dependencies are allowed to vary in space across the 42 locations in the basis grid, represented as points with a color fill that corresponds to the parameter value. The right panel shows the same parameters projected onto the full spatial domain by taking the multiplication $\boldsymbol{\Phi} A_{ij}$ and adjusting the scale as detailed in Section 4.3.2. The color scale is shared for each panel.

values for $\boldsymbol{A}$. The diagonal blocks A_{ii} represent the autocorrelations or temporal dependencies of each variable on themselves. For these blocks, we specify strong positive values. Whereas, for the off-diagonal blocks A_{ij}, $i \neq j$, which represent the temporal cross dependencies of each variable on the other variables, we consider a variety of coefficients and spatial patterns. See Section 4.8.2 for full details on how we constructed each block of $\boldsymbol{A}$.

Using the specified parameter values, we generate the state vectors $\boldsymbol{\alpha}_t$ from the model equations in (4.7). First, $\boldsymbol{\alpha}_0$ is initialized by taking a random draw from a standard normal distribution $N(\mathbf{0}_{MK}, \mathbf{1}_{MK})$. Each subsequent $\boldsymbol{\alpha}_t$ is generated recursively using the transition equation using the above specified values of $\boldsymbol{A}$ and $\boldsymbol{Q}$. To remove the influence of our initial $\boldsymbol{\alpha}_0$ draw on our results, we repeat this process until we have generated 244 $\boldsymbol{\alpha}_t$ vectors and drop the first 100 to obtain the desired $T = 144$ time points. For each $\boldsymbol{\alpha}_t$, the corresponding synthetic observations $\boldsymbol{Y}_t$ are generated directly from the measurement equation with the error terms $\boldsymbol{E}_t$ and $\boldsymbol{H}_t$ generated randomly using covariance matrices $\boldsymbol{V}_t$ and $\boldsymbol{Q}$, respectively.

4.4.2 Simulation Results

To test our MCMC sampler, we fit our MV-STDM to the simulated $\boldsymbol{Y}_t$ data and compare the results to the known values of σ_{it}^2, τ_i^2, and $\boldsymbol{A}$. We specify priors with $\boldsymbol{m}_0 = \mathbf{0}_{MK}$, $\boldsymbol{C}_0 = \boldsymbol{I}_{MK}$, $a_\sigma = b_\sigma = 1$, $a_\tau = b_\tau = 1$, and $\lambda = 1/4$. We produce a single chain of our MCMC of $2,500$ iterations. Burn in was diagnosed within 250 iterations using trace plots, but we dropped the first 500 to obtain a final set of $2,000$ posterior samples.

Table 4.1: Posterior means and 95% credible intervals for parameters estimated from the synthetic dataset. Columns correspond to each variable, denoted by indices $i = 1, 2$, and 3. Results are given for the innovation variance parameter, τ_i^2, as well as the measurement error parameters, σ_{it}^2, at three time points, $t = 1, 72, 144$. The lower and upper bounds of each credible interval are based on the 0.025 and 0.975 quantiles of the posterior samples, respectively.

		$i = 1$		$i = 2$		$i = 3$	
	True	**Post. Mean**	**95% CI**	**Post. Mean**	**95% CI**	**Post. Mean**	**95% CI**
τ_i^2	5	4.89	$(4.67, 5.12)$	4.85	$(4.63, 5.07)$	5.04	$(4.82, 5.27)$
$\sigma_{i,1}^2$	2	2.03	$(1.87, 2.21)$	1.97	$(1.81, 2.14)$	2.00	$(1.84, 2.18)$
$\sigma_{i,72}^2$	2	2.00	$(1.84, 2.17)$	1.97	$(1.81, 2.15)$	2.11	$(1.94, 2.30)$
$\sigma_{i,144}^2$	2	1.99	$(1.84, 2.16)$	2.00	$(1.84, 2.17)$	2.02	$(1.85, 2.19)$

Table 4.1 contains the simulation results for the variance terms τ_i^2 and σ_{it}^2, with posterior means and 95% credible intervals calculated over the 2,000 posterior samples. Overall, the posterior means represent the innovation variance terms τ_i^2 very well. For all three variables, $i = 1, 2, 3$, the true values fall within the credible interval. For the measurement error terms σ_{it}^2, we show the posterior mean estimates and credible intervals for three different time points, $t = 1, 72, 144$, to represent the beginning, middle, and end of the synthetic observations. For all variables, the posterior means are very close to the true parameter value and the posterior credible intervals cover this true value. In addition, we calculated the coverage of the σ_{it}^2 credible intervals over all time points and variables, and obtained a total coverage of 97.22%. This indicates that the σ_{it}^2 parameters are generally recovered well by the model at all time points.

For the transition matrix $\boldsymbol{A}$, Figure 4.3 provides a comparison of the posterior sampling results to the true values. The posterior mean recovers the true values of the coefficients in each A_{ij} block. The only notable difference is some small spatial noise, which is most apparent in blocks such as A_{23} that feature a non-constant spatial pattern. In cases where elements of $\boldsymbol{A}$ are large in magnitude, such as the diagonal blocks $A_{ii}, i = 1, 2, 3$, the credible intervals do not contain 0. This is also seen in cases with weaker temporal dependence, such as A_{21} and A_{32}. For A_{12} and A_{13}, the blocks with 0 temporal dependence, the lower and upper bounds of the credible intervals have a magnitude of about 0.15, demonstrating the width of our credible intervals. For the blocks A_{31} and A_{23} that have nontrivial spatial patterns for the true coefficients, the estimated parameters still match the true values very well. For example, the credible intervals for A_{23} contain only positive values near the north pole, negative values near the south pole, and cover 0 near the horizontal line at the equator, matching the spatial trends in the true coefficient values. Overall, our MCMC sampler does an excellent job at recovering the true values for τ_i^2, σ_{it}^2, and $\boldsymbol{A}$ given the simulated $\boldsymbol{Y}_t$ data.

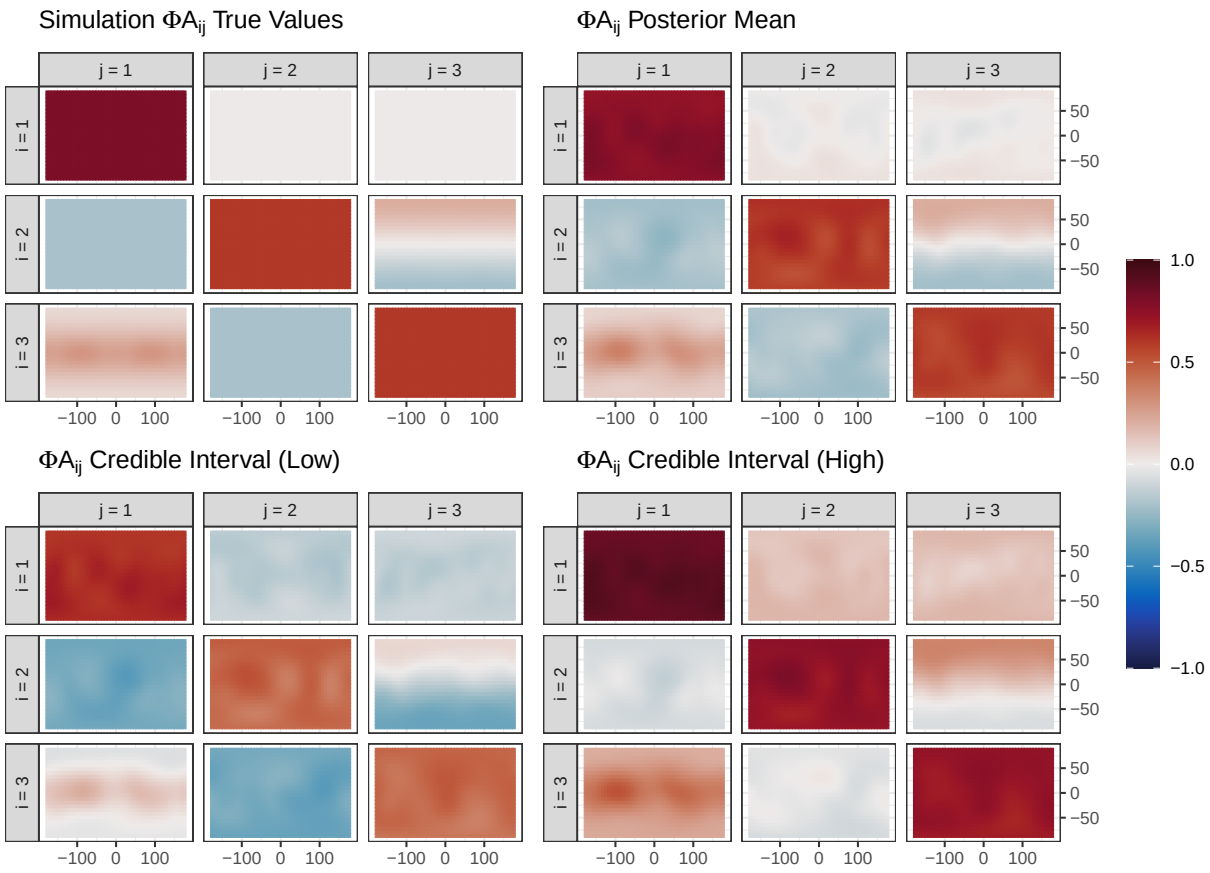


Figure 4.3: Comparison of true $\boldsymbol{A}$ matrix values and posterior sampling results for the synthetic dataset. In all four panels, $\boldsymbol{A}$ is separated into the blocks $\boldsymbol{A}_{ij}$ and multiplied by the basis $\boldsymbol{\Phi}$ to project the results onto the full spatial domain. Posterior means and credible intervals are calculated separately for each element of $\boldsymbol{A}$ over the 2,000 posterior samples. The lower and upper bounds of the credible interval are based on the 0.025 and 0.975 quantiles of the posterior samples, respectively.

4.5 Application

We apply our MV-STDM to study the interaction and evolution of key climate processes following the 1991 Mt. Pinatubo eruption. We focus on three variables from the MERRA-2 Reanalysis dataset: aerosol optical depth (AOD), upwelling longwave radiation (LWR), and 50mb stratospheric temperatures (T50) as described in Section 4.2. It is important to note that we converted each variable to standardized anomalies to remove typical seasonal and spatial trends. First, we discuss the various parameter specifications and prior distributions used in our MCMC sampler to model the MERRA-2 anomalies. After the model specification, our focus is twofold: parameter inference to study the various temporal and multivariate trends present in the data and prediction to test our model's capability in utilizing multivariate spatiotemporal information to interpolate missing observations. For the goal of parameter inference, we start by fitting our MV-STDM to the AOD, LWR, and T50 anomalies from 1984-1995. We study the estimated transition matrix to understand

the evolution of each process during this time period. To assess prediction performance, we compare our MV-STDM to two simpler univariate spatiotemporal approaches. For each model, we hold out a test set of observations over North America in the three years following the eruption.

We consider the MERRA-2 AOD, LWR, and T50 anomalies as our multivariate spatiotemporal fields data with $M = 3$ variables each observed at a common set of $N = 1,152$ spatial locations (on a 24×48 spatial grid) and $T = 144$ time points at a monthly frequency. To fit our MV-STDM in (4.7) to this data, we must specify a few parameters and prior distributions. First, we specify the basis, $\boldsymbol{\Phi}$, by generating $K = 162$ RBFs on an icosahedral grid. To parameterize the spatial autoregression matrix $\boldsymbol{B}$, which is a component of the innovation variance $\boldsymbol{Q}$, we use a value of $\kappa = 2$. To complete our model specification, uninformative priors are assigned according to (4.3.3) with hyperparameters $\boldsymbol{m}_0 = \boldsymbol{0}_{MK}$, $\boldsymbol{C}_0 = \boldsymbol{I}_{MK}$, $a_\sigma = b_\sigma = 1$, $a_\tau = b_\tau = 1$, and $\lambda = 1/4$. Using our MCMC sampler, we computed two chains with a length of $1,500$ iterations each. Computations for each chain took approximately 51 hours. For both chains, satisfactory convergence was diagnosed from trace plots within 250 iterations. We drop the first 500 iterations from each chain and combine the remaining iterations to obtain a final set of $2,000$ posterior samples for each parameter.

4.5.1 Multivariate Model Parameter Inference

In this part of the study, we focus on parameter inference for elements of the transition matrix, $\boldsymbol{A}$, to provide insights into the spatially-varying temporal and multivariate trends captured by our model. To do this, we train the model using historical AOD, LWR, and T50 anomalies from 1984-1995. As discussed in Section 4.3.2, $\boldsymbol{A}$ is separated into various blocks that represent the temporal dependence for one variable on itself or another variable in the previous month. For clarity, we denote anomalies from the previous month as the $t-1$ time step. The posterior mean results for $\boldsymbol{A}$ are provided in Figure 4.4 with each block A_{ij} projected onto the global spatial field using the process detailed in Section 4.3.2. Consider the three blocks on the diagonal, which represent the temporal autocorrelations for the state vectors $\boldsymbol{\alpha}_t^{\text{AOD}}$, $\boldsymbol{\alpha}_t^{\text{LWR}}$, and $\boldsymbol{\alpha}_t^{\text{T50}}$. For all three variables, we see strong positive autocorrelation in almost all regions, especially in the equatorial band between 30°S and 30°N. Across the globe, autocorrelation is strongest for AOD and weaker for LWR and T50. Both LWR and T50 see the strongest temporal self-dependencies near the equator and mid latitude bands (30°N to 60°N and 30°S to 60°S), but these trends taper off, featuring lower but still positive values near the north and south poles. This spatial trend is particularly pronounced for T50. However, there are a few exceptions. First, for AOD, there is one region over northern Africa which exhibits near-zero autocorrelation. This coincides with a neighboring region in central Europe which also exhibits near-zero self dependency for LWR. LWR also features two regions in the southern hemisphere with similarly low autocorrelations.

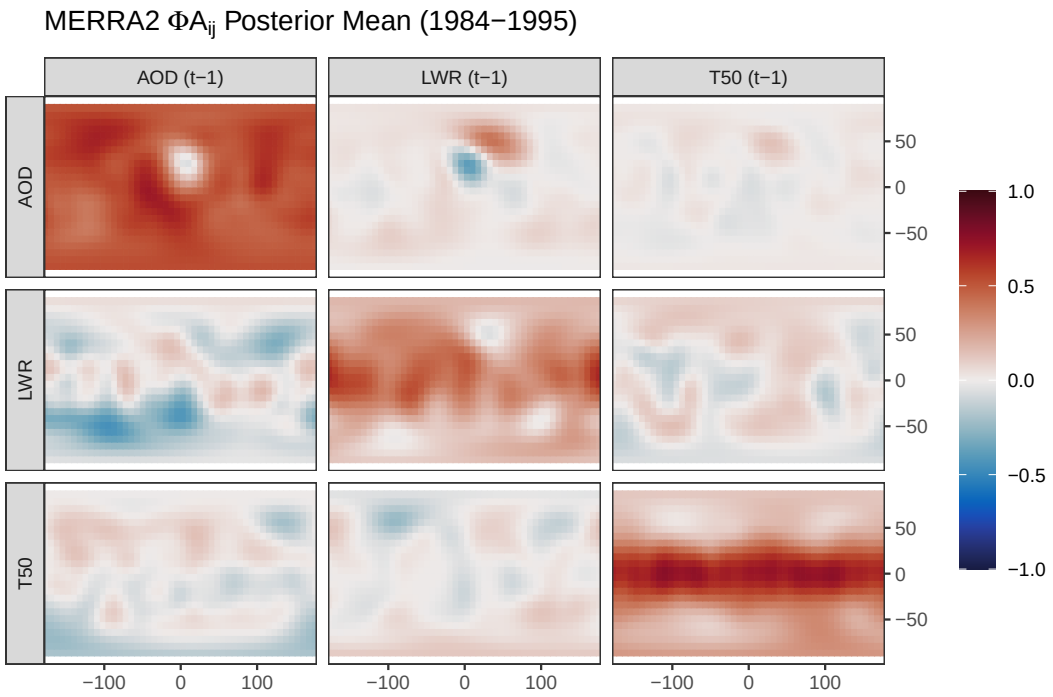


Figure 4.4: Posterior mean of the transition matrix $\boldsymbol{A}$ for the MERRA-2 AOD, LWR, and T50 anomalies. For interpretation, $\boldsymbol{A}$ is separated into blocks A_{ij}, multiplied by the basis $\boldsymbol{\Phi}$ to project the results onto the full spatial domain, and scaled to retain the original magnitude of each coefficient as described in Section 4.3.2. The global spatial domain is represented here by latitude-longitude coordinates on the y-axis and x-axis, respectively. Each of the nine blocks represents the relationship between two variables, with the vertical facets denoting the variable at the current month and the horizontal facets denoting the variable at the previous month, denoted as the $t-1$ time step. Posterior means are calculated separately for each element of $\boldsymbol{A}$ over the 2,000 posterior samples.

Next, we consider the off-diagonal blocks of $\boldsymbol{A}$ which feature the lagged cross-dependencies, or relationships between each variable at time step t and the other two variables at time step $t-1$. First, we focus on the top row of Figure 4.4, where the second and third blocks contain the dependence of $\boldsymbol{\alpha}_t^{\text{AOD}}$ on $\boldsymbol{\alpha}_{t-1}^{\text{LWR}}$ and $\boldsymbol{\alpha}_{t-1}^{\text{T50}}$, respectively. Both LWR and T50 have a near-zero impact on AOD in all regions with one exception. For the lagged cross-dependence of AOD on LWR, the regions over northern Africa and central Europe contain compensating negative and positive dependencies, respectively. These match the regions with abnormally temporal autocorrelations for AOD and LWR discussed in the previous paragraph. AOD seems to feature little or no dependence on previous values of T50 in all regions. Overall, the near-zero cross-dependencies are not surprising in this case because we expect AOD would have an effect on LWR and T50, but not the reverse.

We shift our focus to the second row of Figure 4.4, where the first and third blocks feature the dependence of $\boldsymbol{\alpha}_t^{\text{LWR}}$ on $\boldsymbol{\alpha}_{t-1}^{\text{AOD}}$ and $\boldsymbol{\alpha}_{t-1}^{\text{T50}}$. The impact of AOD on LWR is spatially mixed, with negative values in most regions but pockets of smaller positive values near the equator and north pole. The negative values seen

in many regions make sense given the expected relationship where stratospheric aerosols absorb upwelling longwave radiation, however the positive values near the equator are unexpected. The impact of T50 on LWR is also spatially mixed, with positive and negative values dispersed throughout. However, none of these coefficients are particularly strong, so the impact of T50 on LWR seems generally smaller than the impact of AOD on LWR. This makes sense as changes in AOD are expected to sequentially lead to changes in LWR which in turn can impact T50.

The last row of Figure 4.4 quantifies the dependence of $\boldsymbol{\alpha}_t^{\text{T50}}$ on $\boldsymbol{\alpha}_{t-1}^{\text{AOD}}$ and $\boldsymbol{\alpha}_{t-1}^{\text{LWR}}$ and is of most interest for this application since stratospheric temperatures are the most long lasting impact from the eruption, subsequently leading to significant global surface temperature changes. We notice that AOD has a positive impact on T50 in many regions, especially in the mid-upper latitudes. This matches the expected trend because these regions saw the greatest levels of stratospheric heating following the Mt. Pinatubo eruption as discussed in Section 4.1. However, the magnitude of these coefficients is small, and in some regions (especially near the south pole), we notice a slight negative dependency of T50 on AOD. We see a spatially mixed trend when assessing the dependence of T50 on lagged LWR, with some stronger negative and positive values near the north and south poles, respectively, and near-zero values in the middle latitudes. The magnitude of these coefficients is small across the globe, though there are some notable negative values in the northern hemisphere which may be related to the increased absorption of LWR in that region following the eruption.

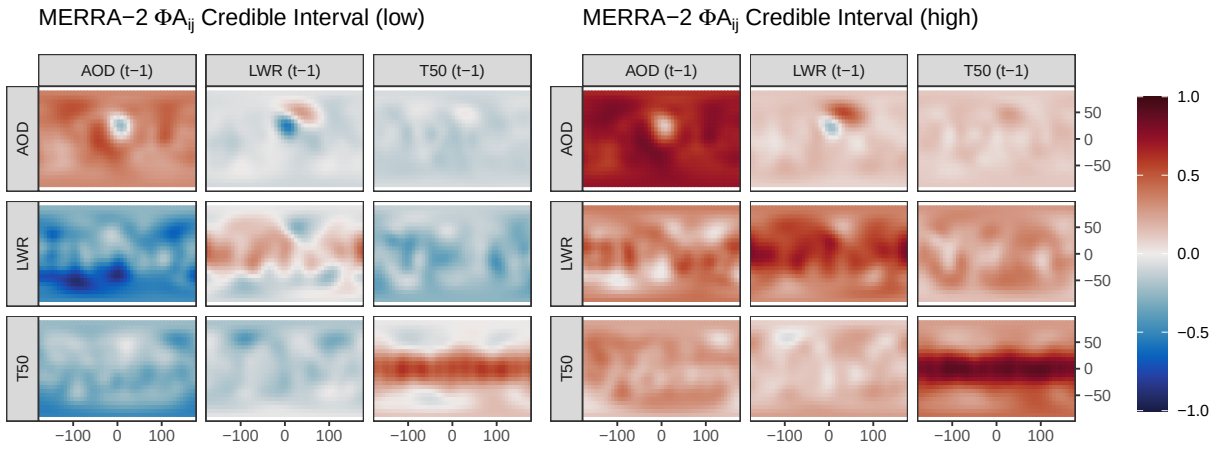


Figure 4.5: 95% posterior credible intervals for the transition matrix, $\boldsymbol{A}$, for the MERRA-2 AOD, LWR, and T50 anomalies. For interpretation, $\boldsymbol{A}$ is separated into blocks A_{ij}, multiplied by the basis $\boldsymbol{\Phi}$ to project the results onto the full spatial domain, and scaled to retain the original magnitude of each coefficient as described in Section 4.3.2. Each of the nine blocks here represents the relationship between one variable and the $t-1$ time step of another variable, and these relationships are allowed to vary over the global, represented here as latitude-longitude coordinates. Credible intervals are calculated separately for each element of $\boldsymbol{A}$ over the 2,000 posterior samples, and the lower and upper bounds of the credible interval are respectively based on the 0.025 and 0.975 quantiles of the posterior samples.

Overall, the posterior mean estimates of the temporal autocorrelations seem strong, but the lagged cross-dependencies exhibit a higher degree of uncertainty. This is reflected in the 95% credible intervals for $\boldsymbol{A}$ shown in Figure 4.5. The credible intervals for the autocorrelations (diagonal blocks) do not contain zero in the regions with strong positive values, but do contain zero in the regions with smaller values. This shows evidence of significant spatially-varying trends in the evolution of each process. For the lagged cross-dependencies, represented by the off-diagonal blocks, the values are relatively close to 0 and the credible intervals do contain 0 in most cases. Notable exceptions include the coefficients representing the impact of LWR on AOD over northern Africa and central Europe, the coefficients representing the impact of AOD on LWR in the southern hemisphere, and the coefficients representing the impact of LWR on T50 in the northern hemisphere. These results suggest that the multivariate trends in the MERRA-2 anomalies may be too small in magnitude to recover with our MV-STDM, or that our model may be overspecified. If the latter is the case, the parameter estimates for $\boldsymbol{A}$ may account for too much of the natural variability in the data, thus obscuring the underlying trends. The simulation results in Section 4.4 indicate that for this data dimension and a similar model specification, the credible intervals produced by our MCMC sampler can be somewhat wide even in the case where the trends in $\boldsymbol{A}$ are clear. So, parameters in $\boldsymbol{A}$ which are small in magnitude could still represent underlying trends in the anomalies. Given that some of the posterior means in the off-diagonal blocks of Figure 4.4 do feature clear spatial trends despite the overall low values, the multivariate model may still contain useful information on the interaction between atmospheric processes.

4.5.2 Assessment of predictive performance

While parameter inference to characterize the evolution of climate processes and understand their interactions is one important focus for our application, we also want to assess our model's capability in predicting downstream impacts of the eruption in localized regions. We test our model's performance in predicting stratospheric temperatures over North America for a three-year period following the eruption (August 1991-July 1994). This region is shown in Figure 4.6, and covers the area bounded by 155°W to 35°W longitude and 5°S to 80°N latitude. This holdout set represents a large missing block of observations in both space and time for a single variable and thus highlights the benefit of our multivariate model. This swath of missing data also mimics patterns of missingness often present in truly observational satellite data such as that collected near-instantaneously by the atmospheric sensor Moderate Resolution Imaging Spectroradiometer [MODIS, Justice et al., 2002] aboard polar orbiting Aqua and Terra satellites. Note that such data did not become available until after 1995 and thus did not measure impacts from the Mt. Pinatubo eruption. However, the interpolation of missing observations is one area in which our model may be useful for understanding the

impact of future eruption events. In particular, good predictive capability can provide early insights into potential near-term impacts after a recently observed event.

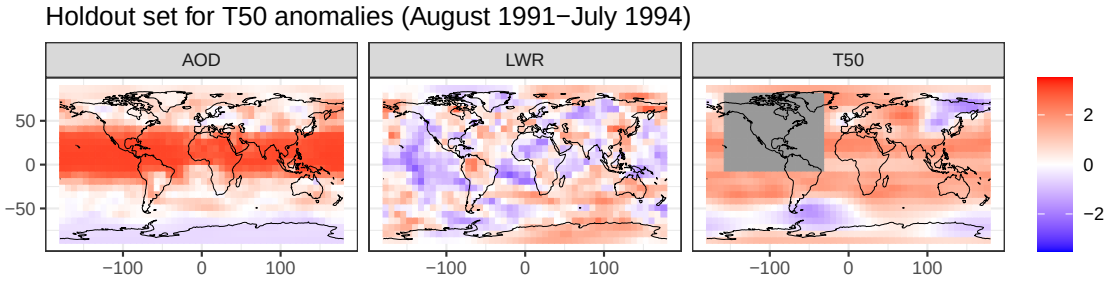
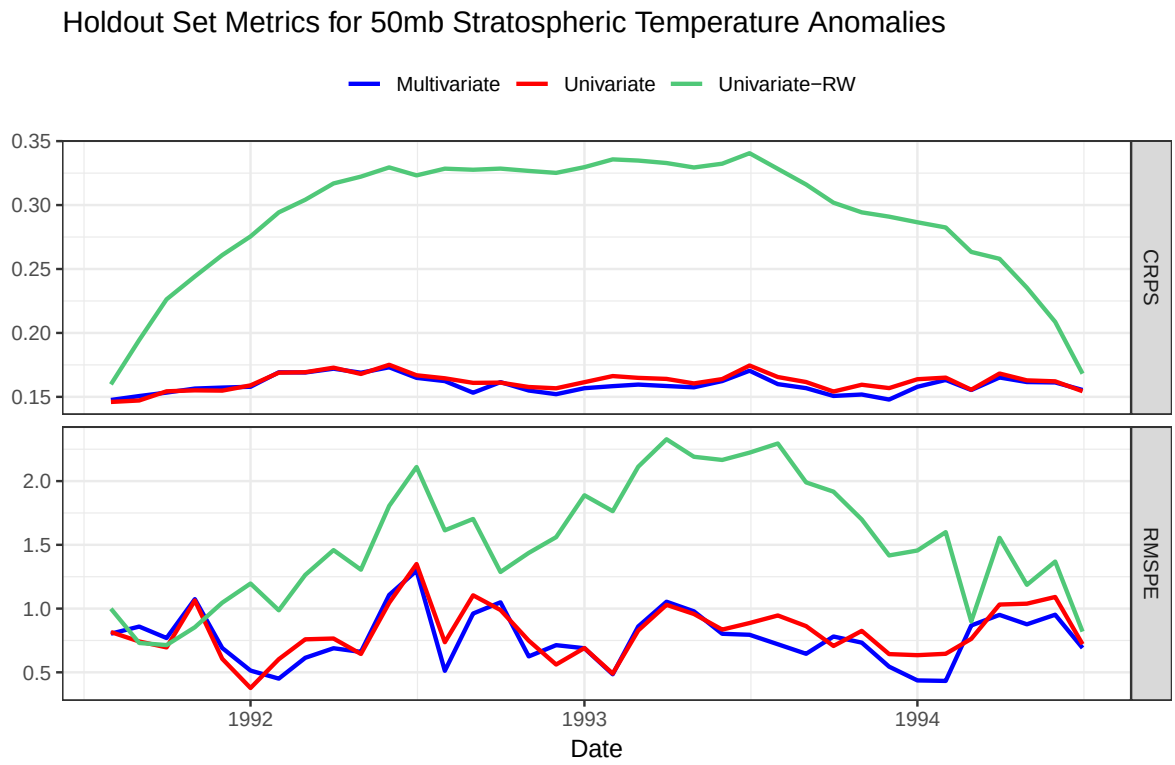


Figure 4.6: Map of the holdout set of observations for stratospheric temperature (T50). The gray region represents the region over which observations are excluded for the August 1991-July 1994 time period when fitting the model. These observations are later used to assess the model's predictive performance. The anomalies shown in the map are for August 1991.

We focus on three submodels for comparison. First, we consider the full MV-STDM discussed in Section 4.5.1 with all three variables (AOD, LWR, T50). Then, to evaluate our model's utility as a (univariate) spatiotemporal dynamic model, we fit a second model with only the T50 anomalies included. While this model only considers temporal dependencies for a single variable, it still offers novel contributions compared to previous approaches due to the spatially-varying structure of the transition matrix. Lastly, we consider a simplified univariate model where we constrain $\boldsymbol{A} = \boldsymbol{I}_K$, again for T50 only. We denote these three models as "Multivariate", "Univariate", and "Univariate-Random Walk", respectively. We call the last model "Univariate-Random Walk" or "Univariate-RW" because it represents a first-order DLM structure [Petris et al., 2009] where the transition equation reduces to $\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t$. This simplified model creates a nonstationary (and thus flexible) temporal structure and eliminates the need to estimate additional model parameters, and is popular for scalable spatiotemporal models [i.e. Finley et al., 2012, Datta et al., 2016, which estimate transition coefficients for the regression parameters, but not the spatial random effect]. For all three models, we follow the specifications discussed at the beginning of Section 4.5 with the exception of $\boldsymbol{m}_0$ and $\boldsymbol{C}_0$, the prior hyperparameters for $\boldsymbol{A}$. For the Multivariate model, these hyperparameters remain the same, but for the Univariate model, we must reduce the dimension of these hyperparameters to match the size of the univariate state vector, so we specify $\boldsymbol{m}_0 = \boldsymbol{0}_K$ and $\boldsymbol{C}_0 = \boldsymbol{I}_K$. For the Univariate-RW model, $\boldsymbol{A}$ is not estimated so the hyperparameters $\boldsymbol{m}_0$ and $\boldsymbol{C}_0$ are not needed. Two chains of 1,500 MCMC iterations are computed for each submodel, and the first 500 iterations are dropped as burn-in to provide a final set of 2,000 posterior samples. In each case, we exclude the data in the holdout set when training the model. Interpolation for the missing observations occurs naturally using the spatial basis function representation in the observation equation.

For each submodel, we use two metrics to assess predictive performance: root mean squared prediction error (RMPSE) and continuous rank probability score [CRPS, Gneiting and Raftery, 2007]. Both metrics are computed based on the posterior sampling results for all observations in the holdout set. RMSPE provides a direct assessment of accuracy based on the mean of the posterior predictive distribution, while CRPS provides a more comprehensive evaluation based on the entire posterior predictive distribution. For each month from August 1991-July 1991, we compute our predictive metrics using all locations in the holdout region to assess the performance of each submodel over time. RMSPE calculations are performed by taking the square root of the average squared prediction error over the entire holdout set, whereas CRPS is first calculated at each location using the empirical distribution function-based approximation [discussed in Krüger et al., 2021] and later averaged over the holdout set. To provide a single number summary, we perform similar calculations of each metric over the entire time period of August 1991-July 1991.



Model	**CRPS** (all months)	**RMSPE** (all months)
Multivariate	0.1593	0.7960
Univariate	0.1618	0.8360
Univariate-RW	0.2905	1.5962

Figure 4.7: Predictive performance for the three submodels on the stratospheric temperature holdout set shown in Figure 4.6. Results are presented for each month from August 1991 through July 1994. The top panel contains the continuous ranked probability score, or CRPS, while the bottom panel contains the root mean squared prediction error, or RMSPE, for each month. Different colors are used to represent each submodel. The table provides averages for each metric over the entire time period.

The results are shown in Figure 4.7. When averaging over the entire time period, our Multivariate model provides the best predictive performance across both metrics. However, our Univariate model is a close second. Interestingly, the simplified Univariate-RW model exhibits much poorer predictive performance with higher values for both metrics. When looking at predictive performance over time, the Multivariate and Univariate models appear to share similar performance each month, especially for CRPS, though the Multivariate model tends to exhibit superior performance most months by a small margin. Whereas, the Univariate-RW model achieves its best performance close to the edges of the holdout time period, where it shares similar RMSPE values and slightly higher CRPS values than the other models, with performance degrading for both metrics in the middle time periods. Our previous results in Section 4.5.1 show significant evidence that the $\boldsymbol{A}$ matrix coefficients representing the autocorrelation of the T50 anomalies are less than 1, suggesting a significant difference in temporal dynamics compared to the Univariate-RW model. These coefficients also vary over space, unlike the Univariate-RW model, allowing the model to adapt to changes in the evolution of the T50 anomalies in different regions. The Univariate-RW model likely struggles with our holdout scenario because it does not share these features. In most cases, our results in Section 4.5.1 suggest that the Multivariate model struggled to capture strong cross-correlations between variables over time, so it is overall not surprising to see that our Univariate model performs almost as well as our Multivariate model. However, the increased predictive performance of the Multivariate model, though slight, does provide evidence that our method can successfully capture and leverage information about the joint evolution of climate processes.

4.6 Discussion

Understanding interactions between atmospheric processes is essential for characterizing the risks and impacts associated with climate change mitigation efforts. To that end, we built a multivariate space-time dynamic model (MV-STDM) which offers a novel parameterization to capture the joint evolution of spatiotemporal processes. We developed a customized MCMC sampler to estimate our model's parameters for multivariate spatiotemporal fields data and established it's effectiveness on a simulated dataset. Finally, we applied our MV-STDM to characterize the joint evolution of aerosol optical depth, longwave radiation, and stratospheric temperatures following the 1991 Mt. Pinatubo eruption. Through parameter inference, we discovered significant spatial trends in the transition matrix and gained insight on the interaction between processes that corroborated known science. In addition, our results demonstrated improved predictive performance when using our proposed MV-STDM when compared to simpler model specifications.

The analysis in Section 4.5.1 shows that our MV-STDM recovers known features of the temporal evolution of AOD, LWR, and T50, such as the strong autocorrelations focused near the equator. However, these

results do not always match our expectations and can exhibit a large degree of uncertainty. For example, the strong interaction between AOD and LWR over northern Africa is likely due to our model overfitting to the anomalies in that region. Because AOD in this region is predominately driven by Saharan dust events, which feature abnormal temporal characteristics in terms of their frequency and duration [Goudie and Middleton, 2001], our time period of 1984-1995 may not be long enough to accurately calculate the climatologies (and therefore anomalies) in this region. Other factors that may impact our results include the increased spatial variability in the longwave radiation anomalies compared to the other variables shown in Figure 4.1 and post-eruption trends such as the decrease in stratospheric temperatures in 1993 due to the Antarctic ozone hole [McCormick et al., 1995]. To alleviate these concerns when using our model, future studies could calculate the climatologies over a longer time period or use strong priors to control the behavior of the transition matrix, i.e., enforcing that AOD has zero dependence on previous values of LWR.

Our model's evolution equation is limited to a linear function of the state vector and considers only one previous time step rather than a longer history. While we made these choices for interpretability and scalability, additional time steps or more complex nonlinear relationships between states could aid in improved spatial predictions. Another option would be to add a "coregionalization" structure to the innovation variance similar to Gelfand et al. [2005]. This would allow for multivariate dependencies in the innovation of the basis coefficients, capturing the contemporaneous, rather than time-lagged, relationship between variables. This may hinder the interpretation of our model as there would be additional sources of multivariate dependencies, but may provide useful insight into the nature of the interactions between observed climate processes.

Compared to some spatial basis function approaches such as Lattice Kriging [Nychka et al., 2015], the number of basis functions considered in our analysis is low. Our MCMC computation time shows that further increases to our basis resolution may not be tractable. So, one important focus for future work is improving the computational speed of our method. To enable high resolution basis representations, Nychka et al. [2015] used a computational approach based on sparse Cholesky decompositions to estimate the basis coefficients. However, the FFBS algorithm used in estimating our $\boldsymbol{\alpha}_t$ coefficients requires solving many linear systems in sequence, making it challenging to maintain the proper data structure to leverage sparse Cholesky decompositions. While we use a spatial model similar to that in Lattice Kriging, we don't estimate the κ parameter in our model as they do. However, estimating this parameter can be computationally expensive, and scaling would be extremely poor due to our expanded focus on multiple variables and time steps as well as our MCMC sampling approach. Further analysis could be done to analyze the sensitivity of our analysis to this parameter or identify separate values for each variable of interest.

Other ideas for improving the scalability of our model include using a different spatial representation of the data or incorporating further approximations. For example, one option would be to consider the spatiotemporal

FFBS approach detailed in Jurek and Katzfuss [2023] which uses a hierarchical Vecchia approximation of Gaussian processes to enable sparse computations. However, it is unclear if this approximation would be suitable for modeling the spatial correlation in our radial basis function coefficients or if a different spatial representation would be required. Alternatively, one way to reduce the dimension of our state vector would be to use a different spatial basis representation. Popular approaches include spectral representations and empirical orthogonal functions (EOFs). However, due to the structure of these bases, each basis function would offer a different global perspective of the spatial field rather than the local perspectives included in our radial basis function representation. So, interpretation of the transition matrix, a key focus of our work, may be more challenging using these representations.

Finally, we see many possibilities for future analyses and applications with our MV-STDM. First, we could estimate our model parameters over different time periods to assess changes in the interaction between AOD, LWR, and T50. Also, we could fit the MV-STDM to output from climate model simulations of each variable. This would allow for an evaluation of the similarity of multivariate interactions between climate models and observational data. By fitting our MV-STDM to multiple runs from the same climate model, we could also assess the impact of internal variability on interactions between climate processes. Similar analyses could be performed to study different impacts of the Mt. Pinatubo eruption, such as impacts on surface temperature or agriculture, or more broadly to study any climate interaction of interest.

4.7 Acknowledgments

This work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, USA, a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration, USA under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

4.8 Supplementary Material

4.8.1 MCMC Sampling Algorithm

Following the model equations in (4.7), we initialize the parameters to some starting values such as

$$\sigma_{it}^2 = 1, \quad \tau_i^2 = 1, \quad \boldsymbol{A} = \boldsymbol{I}_{MK} \quad \boldsymbol{\alpha}_t = \boldsymbol{Z}_t,$$

where $\boldsymbol{Z}_t, t = 1, \ldots, T$ is a $KM \times 1$ matrix with each element drawn independently from a standard normal distribution. We then proceed to sample from the following conditional posterior distributions for each MCMC iteration:

1. For $i = 1, \ldots, M$, $\tau_i^2 \sim IG\left(a_i^*, b_i^*\right)$, where

$$a_i^* = a_\tau + \frac{KT}{2},$$
$$b_i^* = b_\tau + \frac{1}{2}\left(\boldsymbol{\eta}_t^{(i)'} \boldsymbol{B}' \boldsymbol{B} \boldsymbol{\eta}_t^{(i)}\right).$$

2. For $t = 1, \ldots, T$ and $i = 1, \ldots, M$, $\sigma_{it}^2 \sim IG\left(a_{it}^*, b_{it}^*\right)$, where

$$a_{it}^* = a_\sigma + \frac{N}{2},$$
$$b_{it}^* = b_\sigma + \frac{1}{2}\left(\boldsymbol{Y}_t^{(i)} - \boldsymbol{\Phi}\boldsymbol{\alpha}_t^{(i)}\right)'\left(Y_t^{(i)} - \boldsymbol{\Phi}\boldsymbol{\alpha}_t^{(i)}\right).$$

3. Reformulate parameters of the measurement equation as in (4.8), then $\text{vec}(\tilde{\boldsymbol{A}}) \sim N(\boldsymbol{m}_A, \boldsymbol{C}_A)$, where

$$\boldsymbol{C}_A^{-1} = \begin{bmatrix} \boldsymbol{X}_{T-1}^\alpha \\ \vdots \\ \boldsymbol{X}_0^\alpha \end{bmatrix}' \left(\boldsymbol{I}_T \otimes \boldsymbol{Q}^{-1}\right) \begin{bmatrix} \boldsymbol{X}_{T-1}^\alpha \\ \vdots \\ \boldsymbol{X}_0^\alpha \end{bmatrix} + \boldsymbol{V}_0,$$
$$\boldsymbol{m}_A = \boldsymbol{C}_A \left(\begin{bmatrix} \boldsymbol{X}_{T-1}^\alpha \\ \vdots \\ \boldsymbol{X}_0^\alpha \end{bmatrix}' \left(\boldsymbol{I}_T \otimes \boldsymbol{Q}^{-1}\right) \begin{bmatrix} \boldsymbol{\alpha}_T \\ \vdots \\ \boldsymbol{\alpha}_1 \end{bmatrix} + \boldsymbol{V}_0 \boldsymbol{\mu}_0 \right).$$

   Here, $\boldsymbol{\mu}_0$ and $\boldsymbol{V}_0$ are the prior mean and variance for $\text{vec}(\tilde{\boldsymbol{A}})$ following the specification in (4.9).

4. Jointly sample $\boldsymbol{\alpha}_0, \ldots, \boldsymbol{\alpha}_T$ using the FFBS algorithm from Carter and Kohn [1994], Frühwirth-Schnatter [1994], described in Petris et al. [2009] and West and Harrison [1997]. This algorithm handles missing observations in Y_t, and interpolation for these observations occurs naturally in the measurement equation. Our code implementation for this step is based on the code in the 'dlm' R package [Petris, 2010], but includes modifications for the specific structure of our model. First, their algorithm requires computing the Singular Value Decomposition (SVD) of each $\boldsymbol{V}_t$ matrix, which is slow in general but can be completed quickly for diagonal matrices using an analytical form. Second, because our

model specification features a number of sparse matrices, such as $\boldsymbol{\Phi}$, $\boldsymbol{A}$, and $\boldsymbol{B}$, we use sparse matrix multiplication when possible.

4.8.2 Transition Matrix Specification for Synthetic Dataset

Section 4.4.1 discusses the parameter choices for our synthetic data example. Here, we discuss in further detail the parameter specification for $\boldsymbol{A}$, which is depicted in Figure 4.2. For this parameter, we specify different types of relationships for each block A_{ij}, each of which determines the temporal dependence of variable i on variable j. For the autocorrelations, which represent temporal dependencies of each variable on themselves, we specify strong positive values as follows

$$A_{11} = 0.8 * \boldsymbol{I}_K, \qquad A_{22} = A_{33} = 0.6 * \boldsymbol{I}_K.$$

In each case we use a scalar multiple of the identity matrix, so the coefficients are the same for every RBF and thus the pattern of autocorrelation is spatially constant. For the lagged cross correlations, or temporal dependencies of each variable on the other variables, we consider a variety of coefficients and spatial patterns across the different blocks. To test how our model handles spatially constant negative and zero coefficients, we specify

$$A_{12} = A_{13} = 0 * \boldsymbol{I}_K, \qquad A_{21} = A_{32} = -0.2 * \boldsymbol{I}_K.$$

Then, to test spatially-varying dependencies, we introduce a helper function, $\text{lat}(k)$, which provides the latitude coordinate in degrees of the center point of location $k = 1, ..., K$ in the basis grid. For the remaining blocks, we first specify

$$A_{31} = 0.4 * \text{diag}(\tilde{A}_{31}), \qquad \tilde{A}_{31}[k] = 1 - \sqrt{|\text{lat}(k)/90|},$$

where $\tilde{A}_{31}[k]$ denotes the k-th element of the $K \times 1$ vector $\tilde{A}_{31}$. The resulting spatial trend has a value of 0.4 at the equator, but decays to 0 at the north and south poles. Finally, to test a mix of positive and negative dependencies over space, we specify

$$A_{23} = 0.3 * \text{diag}(\tilde{A}_{23}), \qquad \tilde{A}_{23}[k] = \text{lat}(k)/90,$$

where $\tilde{A}_{23}[k]$ denotes the k-th element of the $K \times 1$ vector $\tilde{A}_{23}$. This provides a linear trend starting at 0.3 at the north pole and decreasing to -0.3 at the south pole.

References

Robert Adler, Jian-Jian Wang, Mathew Sapiano, George Huffman, David Bolvin, Eric Nelkin, et al. Global precipitation climatology project (gpcp) climate data record (cdr), version 1.3 (daily). 2020.

JK Angell. Stratospheric warming due to agung, el chichón, and pinatubo taking into account the quasi-biennial oscillation. *Journal of Geophysical Research: Atmospheres*, 102(D8):9479–9485, 1997.

Brian Ayugi, Jiang Zhihong, Huanhuan Zhu, Hamida Ngoma, Hassen Babaousmail, Karim Rizwan, and Victor Dike. Comparison of cmip6 and cmip5 models in simulating mean and extreme precipitation over east africa. *International Journal of Climatology*, 41(15):6474–6496, 2021.

Sudipto Banerjee, Alan Gelfand, Andrew Finley, and Huiyan Sang. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 70: 825–848, 09 2008. doi: 10.1111/j.1467-9868.2008.00663.x.

Sudipto Banerjee, Bradley P Carlin, and Alan E Gelfand. *Hierarchical modeling and analysis for spatial data (2nd Edition)*. Chapman and Hall/CRC, 2014.

Lennart Bengtsson, Stefan Hagemann, and Kevin I Hodges. Can climate trends be calculated from reanalysis data? *Journal of Geophysical Research: Atmospheres*, 109(D11), 2004.

Javier Bernal, Jim Lawrence, Gunay Dogan, and Robert Hagwood. On computing elastic shape distances between curves in d-dimensional space. *NIST Technical Note*, 2164, 2021.

H Nicholas Bhalme and D Albert Mooley. Large-scale droughts/floods and monsoon circulation. 1980.

Clément Bonet, Paul Berg, Nicolas Courty, François Septier, Lucas Drumetz, and Minh-Tan Pham. Spherical sliced-wasserstein. *arXiv preprint arXiv:2206.08780*, 2022.

Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and radon wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51:22–45, 2015.

Alberto Boretti. Reassessing the cooling that followed the 1991 volcanic eruption of mt. pinatubo. *Journal of Atmospheric and Solar-Terrestrial Physics*, 256:106187, 2024. ISSN 1364-6826. doi: https://doi.org/10.1016/j.jastp.2024.106187. URL https://www.sciencedirect.com/science/article/pii/S1364682624000154.

Jonathan R. Bradley, Scott H. Holan, and Christopher K. Wikle. Multivariate spatio-temporal models for high-dimensional areal data with application to Longitudinal Employer-Household Dynamics. *The Annals of Applied Statistics*, 9(4):1761 – 1791, 2015.

William M Briggs and Richard A Levine. Wavelets and field forecast verification. *Monthly Weather Review*, 125(6):1329–1341, 1997.

Federico A Bugni and Joel L Horowitz. Permutation tests for equality of distributions of functional data. *Journal of Applied Econometrics*, 36(7):861–877, 2021.

Chris K Carter and Robert Kohn. On Gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

Joshua CC Chan. *Large Bayesian vector autoregressions*. Springer, 2020.

Chao-An Chen, Huang-Hsiung Hsu, and Hsin-Chien Liang. Evaluation and comparison of cmip6 and cmip5 model performance in simulating the seasonal extreme precipitation in the western north pacific and east asia. *Weather and Climate Extremes*, 31:100303, 2021.

Noel Cressie and G Johannesson. Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70:209–226, 2008.

Noel Cressie and Christopher Wikle. Hierarchical dynamical spatio-temporal models. In Jan Fagerberg, David C. Mowery, and Richard R. Nelson, editors, *Statistics for Spatio-Temporal Data*, chapter 7, pages 361–440. Wiley, Hoboken, NJ, 2011.

Noel Cressie, Martina Pavlicová, and Thomas J Santner. Detecting signals in fmri data using powerful fdr procedures. *Statistics and its interface*, 1(1):23–32, 2008.

Noel Cressie, Matthew Sainsbury-Dale, and Andrew Zammit-Mangion. Basis-Function Models in Spatial Statistics. *Annual Review of Statistics and Its Application*, 9(1):373–400, 2022. doi: http://dx.doi.org/10.1146/annurev-statistics-040120-020733. URL https://ssrn.com/abstract=4065357.

Li Cui, Xin Qi, Chengfeng Wen, Na Lei, Xinyuan Li, Min Zhang, and Xianfeng Gu. Spherical optimal transportation. *Computer-Aided Design*, 115:181–193, 2019.

Abhirup Datta, Sudipto Banerjee, Andrew O Finley, and Alan E Gelfand. Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812, 2016.

Ishan Deshpande, Yuan-Ting Hu, Ruoyu Sun, Ayis Pyrros, Nasir Siddiqui, Sanmi Koyejo, Zhizhen Zhao, David Forsyth, and Alexander G Schwing. Max-sliced wasserstein distance and its use for gans. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10648–10656, 2019.

James R Driscoll and Dennis M Healy. Computing fourier transforms and convolutions on the 2-sphere. *Advances in applied mathematics*, 15(2):202–250, 1994.

Yi Du, Dagang Wang, Jinxin Zhu, Dayang Wang, Xiaoxing Qi, and Jingheng Cai. Comprehensive assessment of cmip5 and cmip6 models in simulating and projecting precipitation over the global land. *International Journal of Climatology*, 42(13):6859–6875, 2022.

Dirk Eddelbuettel and Romain François. Rcpp: Seamless r and c++ integration. *Journal of statistical software*, 40:1–18, 2011.

Dirk Eddelbuettel and Conrad Sanderson. Rcpparmadillo: Accelerating r with high-performance c++ linear algebra. *Computational statistics & data analysis*, 71:1054–1063, 2014.

V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor. Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958, 2016a. doi: 10.5194/gmd-9-1937-2016. URL https://gmd.copernicus.org/articles/9/1937/2016/.

Veronika Eyring, Sandrine Bony, Gerald A Meehl, Catherine A Senior, Bjorn Stevens, Ronald J Stouffer, and Karl E Taylor. Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958, 2016b.

Veronika Eyring, Peter M Cox, Gregory M Flato, Peter J Gleckler, Gab Abramowitz, Peter Caldwell, William D Collins, Bettina K Gier, Alex D Hall, Forrest M Hoffman, et al. Taking climate model evaluation to the next level. *Nature Climate Change*, 9(2):102–110, 2019.

John Fasullo and PJ Webster. A hydrological definition of indian monsoon onset and withdrawal. *Journal of Climate*, 16(19):3200–3211, 2003.

Andrew O. Finley, Huiyan Sang, Sudipto Banerjee, and Alan E. Gelfand. Improving the performance of predictive process modeling for large datasets. *Computational Statistics and Data Analysis*, 53(8): 2873–2884, 2009. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2008.09.008. URL https://www.sciencedirect.com/science/article/pii/S0167947308004386.

Andrew O Finley, Sudipto Banerjee, and Alan E Gelfand. Bayesian dynamic modeling for large space-time datasets using Gaussian predictive processes. *Journal of Geographical Systems*, 14:29–47, 2012.

Andrew O. Finley, Abhirup Datta, Bruce D. Cook, Douglas C. Morton, Hans E. Andersen, and Sudipto Banerjee. Efficient algorithms for bayesian nearest neighbor gaussian processes. *Journal of Computational and Graphical Statistics*, 28(2):401–414, 2018.

Gregory Flato, Jochem Marotzke, Babatunde Abiodun, Pascale Braconnot, Sin Chan Chou, William Collins, Peter Cox, Fatima Driouech, Seita Emori, Veronika Eyring, et al. Evaluation of climate models. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 741–866. Cambridge University Press, 2014.

Andrew R Friedman, Yen-Ting Hwang, John CH Chiang, and Dargan MW Frierson. Interhemispheric temperature asymmetry over the twentieth century and in future projections. *Journal of Climate*, 26(15): 5419–5433, 2013.

Sylvia Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15(2):183–202, 1994. doi: https://doi.org/10.1111/j.1467-9892.1994.tb00184.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9892.1994.tb00184.x.

Ronald Gelaro, Will McCarty, Max J Suárez, Ricardo Todling, Andrea Molod, Lawrence Takacs, Cynthia A Randles, Anton Darmenov, Michael G Bosilovich, Rolf Reichle, et al. The modern-era retrospective analysis for research and applications, version 2 (merra-2). *Journal of climate*, 30(14):5419–5454, 2017.

Alan E. Gelfand. *Multivariate Spatial Process Models*, pages 1985–2016. Springer Berlin Heidelberg, 2021. ISBN 978-3-662-60723-7. doi: 10.1007/978-3-662-60723-7_120.

Alan E. Gelfand, Sudipto Banerjee, and Dani Gamerman. Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics*, 16(5):465–479, 2005. doi: https://doi.org/10.1002/env.715. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/env.715.

Irène Gijbels and Stanislav Nagy. On a general definition of depth for functional data. 2017.

Peter Gleckler, Charles Doutriaux, Paul Durack, Karl Taylor, Yuying Zhang, Dean Williams, Erik Mason, and Jérôme Servonnat. A more powerful reality test for climate models. *Eos*, 97, 2016.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.

Andrew S Goudie and Nick J Middleton. Saharan dust storms: nature and consequences. *Earth-science reviews*, 56(1-4):179–204, 2001.

Xiaohan Guo, Sebastian Kurtek, and Karthik Bharath. Variograms for kriging and clustering of spatial functional data with phase variation. *Spatial statistics*, 51:100687, 2022.

Kyung-Ja Ha, Suyeon Moon, Axel Timmermann, and Daeha Kim. Future changes of summer monsoon characteristics and evaporative demand over asia in cmip6 simulations. *Geophysical Research Letters*, 47 (8):e2020GL087492, 2020.

Peter Hall and Ingrid Van Keilegom. Two-sample tests in functional data analysis starting from discrete data. *Statistica Sinica*, pages 1511–1531, 2007.

James Hansen, Makiko Sato, and Reto Ruedy. Perception of climate change. *Proceedings of the National Academy of Sciences*, 109(37):E2415–E2423, 2012.

Trevor Harris, Bo Li, Nathan J Steiger, Jason E Smerdon, Naveen Narisetty, and J Derek Tucker. Evaluating proxy influence in assimilated paleoclimate reconstructions—testing the exchangeability of two ensembles of spatial processes. *Journal of the American Statistical Association*, 116(535):1100–1113, 2021a.

Trevor Harris, J Derek Tucker, Bo Li, and Lyndsay Shand. Elastic depths for detecting shape anomalies in functional data. *Technometrics*, 63(4):466–476, 2021b.

Emmanuel Hartman, Yashil Sukurdeep, Nicolas Charon, Eric Klassen, and Martin Bauer. Supervised deep learning of elastic srv distances on the shape space of curves. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4425–4433, 2021.

Matthew J Heaton, Abhirup Datta, Andrew O Finley, Reinhard Furrer, Joseph Guinness, Rajarshi Guhaniyogi, Florian Gerber, Robert B Gramacy, Dorit Hammerling, Matthias Katzfuss, et al. A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, 24:398–425, 2019.

MJ Heaton, M Katzfuss, C Berrett, and DW Nychka. Constructing valid spatial processes on the sphere using kernel convolutions. *Environmetrics*, 25(1):2–15, 2014.

Amanda S Hering and Marc G Genton. Comparing spatial predictions. *Technometrics*, 53(4):414–425, 2011.

H. Hersbach, B. Bell, P. Berrisford, G. Biavati, A. Horányi, J. Muñoz Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, and J-N. Thépaut. Era5 monthly averaged data on single levels from 1940 to present, 2023. Accessed on 02-03-2021.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.

Lajos Horváth, Piotr Kokoszka, and Ron Reeder. Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1): 103–122, 2013.

Frédéric Hourdin, Thorsten Mauritsen, Andrew Gettelman, Jean-Christophe Golaz, Venkatramani Balaji, Qingyun Duan, Doris Folini, Duoying Ji, Daniel Klocke, Yun Qian, et al. The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, 98(3):589–602, 2017.

Simon Hubbert and Janin Jäger. Generalised wendland functions for the sphere. *Advances in Computational Mathematics*, 49(1):3, 2023.

Gabriel Huerta, Bruno Sansó, and Jonathan R. Stroud. A Spatiotemporal Model for Mexico City Ozone Levels. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 53(2):231–248, 03 2004. ISSN 0035-9254. doi: 10.1046/j.1467-9876.2003.05100.x. URL https://doi.org/10.1046/j.1467-9876.2003.05100.x.

George J Huffman, Robert F Adler, Mark M Morrissey, David T Bolvin, Scott Curtis, Robert Joyce, Brad McGavock, and Joel Susskind. Global precipitation at one-degree daily resolution from multisatellite observations. *Journal of hydrometeorology*, 2(1):36–50, 2001.

Jacek Jachymski and Jakub Klima. Around perov's fixed point theorem for mappings on generalized metric spaces. *Fixed Point Theory*, 17(2), 2016.

John E Janowiak, Arnold Gruber, CR Kondragunta, Robert E Livezey, and George J Huffman. A comparison of the ncep–ncar reanalysis precipitation and the gpcp rain gauge–satellite combined dataset with observational error considerations. *Journal of Climate*, 11(11):2960–2979, 1998.

Shantanu H Joshi, Eric Klassen, Anuj Srivastava, and Ian Jermyn. A novel representation for riemannian analysis of elastic curves in rn. In *2007 IEEE conference on computer vision and pattern recognition*, pages 1–7. IEEE, 2007.

Marcin Jurek and Matthias Katzfuss. Scalable spatio-temporal smoothing via hierarchical sparse Cholesky decomposition. *Environmetrics*, 34(1):e2757, 2023. doi: https://doi.org/10.1002/env.2757. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2757.

CO Justice, JRG Townshend, EF Vermote, E Masuoka, RE Wolfe, Nazmi Saleous, DP Roy, and JT Morisette. An overview of modis land data processing and product status. *Remote sensing of Environment*, 83(1-2): 3–15, 2002.

Masao Kanamitsu, Wesley Ebisuzaki, Jack Woollen, Shi-Keng Yang, JJ Hnilo, M Fiorino, and GL Potter. Ncep–doe amip-ii reanalysis (r-2). *Bulletin of the American Meteorological Society*, 83(11):1631–1644, 2002.

Rizwan Karim, Guirong Tan, Brian Ayugi, Hassen Babaousmail, and Fei Liu. Evaluation of historical cmip6 model simulations of seasonal mean temperature over pakistan during 1970–2014. *Atmosphere*, 11(9):1005, 2020.

Arie Kattenberg, F Giorgi, Hartmut Grassl, GA Meehl, JFB Mitchell, RJ Stouffer, T Tokioka, AJ Weaver, and TML Wigley. Climate models: projections of future climate. In *Climate Change 1995: the science of climate change. Contribution of WG1 to the Second Assessment Report of the IPCC*, pages 299–357. Cambridge University Press, 1996.

Anja Katzenberger, Jacob Schewe, Julia Pongratz, and Anders Levermann. Robust increase of indian monsoon rainfall and its variability under future warming in cmip6 models. *Earth System Dynamics*, 12(2):367–386, 2021.

Matthias Katzfuss and Noel Cressie. Bayesian hierarchical spatio-temporal smoothing for very large datasets. *Environmetrics*, 23(1):94–107, 2012.

Matthias Katzfuss and Wenlong Gong. A class of multi-resolution approximations for large spatial datasets. *Statistica Sinica*, 01 2020. doi: 10.5705/ss.202018.0285.

Matthias Katzfuss and Joseph Guinness. A General Framework for Vecchia Approximations of Gaussian Processes. *Statistical Science*, 36(1):124 – 141, 2021. doi: 10.1214/19-STS755. URL https://doi.org/10.1214/19-STS755.

Maxwell Kelley, Gavin A Schmidt, Larissa S Nazarenko, Susanne E Bauer, Reto Ruedy, Gary L Russell, Andrew S Ackerman, Igor Aleinov, Michael Bauer, Rainer Bleck, et al. Giss-e2. 1: Configurations and climatology. *Journal of Advances in Modeling Earth Systems*, 12(8):e2019MS002025, 2020.

John T Kent. The fisher-bingham distribution on the sphere. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(1):71–80, 1982.

Dibesh Khadka, Mukand S Babel, Abayomi A Abatan, and Matthew Collins. An evaluation of cmip5 and cmip6 climate models in simulating summer rainfall in the southeast asian monsoon domain. *International Journal of Climatology*, 42(2):1181–1202, 2022.

Jong Uhn Kim. Invariant measures for a stochastic nonlinear schrödinger equation. *Indiana University mathematics journal*, pages 687–717, 2006.

William Kleiber, Douglas Nychka, and Soutir Bandyopadhyay. A model for large multivariate spatial data sets. *Statistica Sinica*, 29(3):1085–1104, 2019. ISSN 10170405, 19968507. URL https://www.jstor.org/stable/26705994.

Soheil Kolouri, Kimia Nadjahi, Umut Simsekli, Roland Badeau, and Gustavo Rohde. Generalized sliced wasserstein distances. *Advances in neural information processing systems*, 32, 2019.

Gopinadh Konda and Naresh Krishna Vissa. Evaluation of cmip6 models for simulations of surplus/deficit summer monsoon conditions over india. *Climate Dynamics*, 60(3):1023–1042, 2023.

Fabian Krüger, Sebastian Lerch, Thordis Thorarinsdottir, and Tilmann Gneiting. Predictive inference based on markov chain monte carlo output. *International Statistical Review*, 89(2):274–301, 2021.

Adam AL Levy, Mark Jenkinson, William Ingram, and Myles Allen. Correcting precipitation feature location in general circulation models. *Journal of Geophysical Research: Atmospheres*, 119(23):13–350, 2014.

Bo Li and Jason E Smerdon. Defining spatial comparison metrics for evaluation of paleoclimatic field reconstructions of the common era. *Environmetrics*, 23(5):394–406, 2012.

Bo Li, Xianyang Zhang, and Jason E Smerdon. Comparison between spatio-temporal random processes and application to climate model data. *Environmetrics*, 27(5):267–279, 2016.

Jingjing Li, Ran Huo, Hua Chen, Ying Zhao, and Tianhui Zhao. Comparative assessment and future prediction using cmip6 and cmip5 for annual precipitation and extreme precipitation simulation. *Frontiers in Earth Science*, 9, 2021. ISSN 2296-6463. doi: 10.3389/feart.2021.687976. URL https://www.frontiersin.org/article/10.3389/feart.2021.687976.

Ruonan Li, Pavan Turaga, Anuj Srivastava, and Rama Chellappa. Differential geometric representations and algorithms for some pattern recognition and computer vision problems. *Pattern Recognition Letters*, 43: 3–16, 2014.

Tao Li and Jinwen Ma. Functional data clustering analysis via the learning of gaussian processes with wasserstein distance. In *Neural Information Processing: 27th International Conference, ICONIP 2020, Bangkok, Thailand, November 23–27, 2020, Proceedings, Part II 27*, pages 393–403. Springer, 2020.

Zhengyang Li, Tie Liu, Yue Huang, Jiabin Peng, and Yunan Ling. Evaluation of the cmip6 precipitation simulations over global land. *Earth's Future*, 10(8):e2021EF002500, 2022.

Robert Lund and Bo Li. Revisiting climate region definitions via clustering. *Journal of Climate*, 22(7): 1787–1800, 2009.

H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer Science & Business Media, 2005. URL https://link.springer.com/book/10.1007/978-3-540-27752-1.

M McCormick, Thomason L, and C Trepte. Atmopsheric effects of the mt pinatubo eruption. *Nature*, 373: 399–404, 1995. doi: https://doi.org/10.1038/373399a0.

Carlos R Mechoso, Andrew W Robertson, N Barth, MK Davey, P Delecluse, PR Gent, S Ineson, B Kirtman, Mojib Latif, H Le Treut, et al. The seasonal cycle over the tropical pacific in coupled ocean–atmosphere general circulation models. *Monthly Weather Review*, 123(9):2825–2838, 1995.

Vasubandhu Misra, Amit Bhardwaj, and Akhilesh Mishra. Local onset and demise of the indian summer monsoon. *Climate Dynamics*, 51:1609–1622, 2018.

Hamida Ngoma, Wang Wen, Brian Ayugi, Hassen Babaousmail, Rizwan Karim, and Victor Ongoma. Evaluation of precipitation simulations in cmip6 models over uganda. *International Journal of Climatology*, 41(9):4743–4768, 2021.

Khai Nguyen and Nhat Ho. Revisiting sliced wasserstein on images: From vectorization to convolution. *Advances in Neural Information Processing Systems*, 35:17788–17801, 2022.

Khai Nguyen and Nhat Ho. Energy-based sliced wasserstein distance. *Advances in Neural Information Processing Systems*, 36, 2024.

Douglas Nychka, Soutir Bandyopadhyay, Dorit Hammerling, Finn Lindgren, and Stephan Sain. A multiresolution gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics*, 24(2):579–599, 2015.

Douglas Nychka, Dorit Hammerling, Stephan Sain, and Nathan Lenssen. LatticeKrig: Multiresolution Kriging Based on Markov Random Fields, 2016. URL https://github.com/NCAR/LatticeKrig. R package version 8.4.

Giovanni Petris. An r package for dynamic linear models. *Journal of statistical software*, 36:1–16, 2010.

Giovanni Petris, Sonia Petrone, and Patrizia Campagnoli. *Dynamic linear models with R*. Springer Science and Business Media, 2009. URL https://link.springer.com/book/10.1007/b135794.

Gina-Maria Pomann, Ana-Maria Staicu, and Sujit Ghosh. A two sample distribution-free test for functional data with application to a diffusion tensor imaging study of multiple sclerosis. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, 65(3):395, 2016.

Emilio Porcu, Moreno Bevilacqua, and Marc G Genton. Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. *Journal of the American Statistical Association*, 111(514): 888–898, 2016.

Venkatraman Prasanna. Impact of monsoon rainfall on the total foodgrain yield over india. *Journal of earth system science*, 123(5):1129–1145, 2014.

Michael Quellmalz, Robert Beinert, and Gabriele Steidl. Sliced optimal transport on the sphere. *Inverse Problems*, 39(10):105005, 2023.

Jouni Raäisaänen. How reliable are climate models? *Tellus A: Dynamic Meteorology and Oceanography*, 59 (1):2–29, 2007.

Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 435–446. Springer, 2011.

TV Ramachandra and BV Shruthi. Spatial mapping of renewable energy potential. *Renewable and sustainable energy reviews*, 11(7):1460–1480, 2007.

James O Ramsay and Bernard W Silverman. *Fitting differential equations to functional data: Principal differential analysis*. Springer, 2005.

David A Randall, Richard A Wood, Sandrine Bony, Robert Colman, Thierry Fichefet, John Fyfe, Vladimir Kattsov, Andrew Pitman, Jagadish Shukla, Jayaraman Srinivasan, et al. Climate models and their evaluation. In *Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC (FAR)*, pages 589–662. Cambridge University Press, 2007.

S Sambasiva Rao. On vector valued metric spaces. *Palestine Journal of Mathematics*, 4(2):416–418, 2015.

Philip J Rasch, Simone Tilmes, Richard P Turco, Alan Robock, Luke Oman, Chih-Chieh (Jack) Chen, Georgiy L Stenchikov, and Rolando R Garcia. An overview of geoengineering of climate using stratospheric sulphate aerosols. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 366(1882):4007–4037, 2008. doi: 10.1098/rsta.2008.0131.

Alan Robock. Volcanic eruptions and climate. *Reviews of Geophysics*, 38(2):191–219, 2000. doi: https://doi.org/10.1029/1998RG000054. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/1998RG000054.

Alan Robock, Allison Marquardt, Ben Kravitz, and Georgiy Stenchikov. Benefits, risks, and costs of stratospheric geoengineering. *Geophysical Research Letters*, 36(19), 2009. doi: https://doi.org/10.1029/2009GL039209. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009GL039209.

Rémy Roca, Ziad S Haddad, Fumie F Akimoto, Lisa Alexander, Ali Behrangi, George Huffman, Seiji Kato, Chris Kidd, Pierre-Emmanuel Kirstetter, Takuji Kubota, et al. The joint ipwg/gewex precipitation assessment, 2021.

Richard B. Rood. *Validation of Climate Models: An Essential Practice*, pages 737–762. Springer International Publishing, Cham, 2019. ISBN 978-3-319-70766-2. doi: 10.1007/978-3-319-70766-2_30. URL https://doi.org/10.1007/978-3-319-70766-2_30.

Jonathan Rougier, Aoibheann Brady, Jonathan Bamber, Stephen Chuter, Sam Royston, Bramha Dutt Vishwakarma, Richard Westaway, and Yann Ziegler. The scope of the Kalman filter for spatio-temporal applications in environmental science. *Environmetrics*, 34(1):e2773, 2023. doi: https://doi.org/10.1002/env.2773. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2773.

Matthew Sainsbury-Dale, Andrew Zammit-Mangion, and Noel Cressie. Modeling big, heterogeneous, non-gaussian spatial and spatio-temporal data using frk. *Journal of Statistical Software*, 108(10):1–39, 2024. doi: 10.18637/jss.v108.i10. URL https://www.jstatsoft.org/index.php/jss/article/view/v108i10.

KPR Sastry, GA Naidu, Tadesse Bekeshie, and MA Rahamatulla. A common fixed point theorem for four self maps in complex valued and vector valued metric spaces. *Int. J. Math. Arc*, 3:2680–2685, 2012.

Pierre Sepulchre, Arnaud Caubel, Jean-Baptiste Ladant, Laurent Bopp, Olivier Boucher, Pascale Braconnot, Patrick Brockmann, Anne Cozic, Yannick Donnadieu, Jean-Louis Dufresne, et al. Ipsl-cm5a2–an earth system model designed for multi-millennial climate simulations. *Geoscientific Model Development*, 13(7):3011–3053, 2020.

Xiaotong Shen, Hsin-Cheng Huang, and Noel Cressie. Nonparametric hypothesis testing for a spatial signal. *Journal of the American Statistical Association*, 97(460):1122–1140, 2002.

Seth E Snell, Sucharita Gopal, and Robert K Kaufmann. Spatial interpolation of surface air temperatures using artificial neural networks: Evaluating their use for downscaling gcms. *Journal of Climate*, 13(5): 886–895, 2000.

Anuj Srivastava and Eric P Klassen. *Functional and shape data analysis*, volume 1. Springer, 2016.

Anuj Srivastava, Wei Wu, Sebastian Kurtek, Eric Klassen, and James Stephen Marron. Registration of functional data using fisher-rao metric. *arXiv preprint arXiv:1103.3817*, 2011.

Ana-Maria Staicu, Yingxing Li, Ciprian M Crainiceanu, and David Ruppert. Likelihood ratio tests for dependent data with applications to longitudinal and functional data analysis. *Scandinavian Journal of Statistics*, 41(4):932–949, 2014.

Elias M Stein and Rami Shakarchi. *Functional analysis: introduction to further topics in analysis*, volume 4. Princeton University Press, 2011.

Jonathan R. Stroud, Peter Muller, and Bruno Sanso. Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 63(4):673–689, 2001. ISSN 13697412, 14679868.

Chris Sutherland, Angela K Fuller, and J Andrew Royle. Modelling non-euclidean movement and landscape connectivity in highly structured ecological networks. *Methods in Ecology and Evolution*, 6(2):169–177, 2015.

István Szapudi. Introduction to higher order spatial statistics in cosmology. In *Data Analysis in Cosmology*, pages 457–492. Springer, 2008.

FJ Tapiador, A Navarro, V Levizzani, E García-Ortega, GJ Huffman, C Kidd, PA Kucera, CD Kummerow, H Masunaga, WA Petersen, et al. Global precipitation measurements for validating climate models. *Atmospheric Research*, 197:1–20, 2017.

Ryan J Tibshirani. *The solution path of the generalized lasso*. Stanford University, 2011.

Ryan J Tibshirani. Adaptive piecewise polynomial estimation via trend filtering. 2014.

S Tonellato. Bayesian dynamic linear models for spatial time series. *Rapporto di Ricerca 5/1997*, 1997.

Kevin E Trenberth. Changes in precipitation with climate change. *Climate research*, 47(1-2):123–138, 2011.

J. Derek Tucker. *fdasrvf: Elastic Functional Data Analysis*, 2020. URL https://CRAN.R-project.org/package=fdasrvf. R package version 1.9.3.

J Derek Tucker and Drew Yarger. Elastic functional changepoint detection of climate impacts from localized sources. *Environmetrics*, 35(1):e2826, 2024.

J. Derek Tucker, Wei Wu, and Anuj Srivastava. Generative models for functional data using phase and amplitude separation. *Computational Statistics Data Analysis*, 61:50–66, 2013. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2012.12.001. URL https://www.sciencedirect.com/science/article/pii/S0167947312004227.

J Derek Tucker, Matthew T Martinez, and Jose M Laborde. Dimensionality reduction using elastic measures. *arXiv preprint arXiv:2209.04933*, 2022.

SS Vallender. Calculation of the wasserstein distance between probability distributions on the line. *Theory of Probability & Its Applications*, 18(4):784–786, 1974.

A. V. Vecchia. Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 297–312, 1988.

Cédric Villani. The wasserstein distances. *Optimal Transport: Old and New*, pages 93–111, 2009.

Gabriele Vissio, Valerio Lembo, Valerio Lucarini, and Michael Ghil. Evaluating the performance of climate models based on wasserstein distance. *Geophysical Research Letters*, 47(21):e2020GL089385, 2020.

Bin Wang, Sulochana Gadgil, and K Rupa Kumar. The asian monsoon—agriculture and economy. *The Asian Monsoon*, pages 651–683, 2006.

Bin Wang et al. Rainy season of the asian–pacific summer monsoon. *Journal of Climate*, 15(4):386–398, 2002.

Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Müller. Functional data analysis. *Annual Review of Statistics and its application*, 3:257–295, 2016.

Weile Wang, Bruce T Anderson, Dara Entekhabi, Dong Huang, Yin Su, Robert K Kaufmann, and Ranga B Myneni. Intraseasonal interactions between temperature and vegetation over the boreal forests. *Earth Interactions*, 11(18):1–30, 2007.

Warren M Washington and Claire Parkinson. *Introduction to three-dimensional climate modeling*. University science books, 2005.

Holger Wendland. Error estimates for interpolation by compactly supported radial basis functions of minimal degree. *Journal of approximation theory*, 93(2):258–272, 1998.

Mike West and Jeff Harrison. *Bayesian Forecasting and Dynamic Models*. Springer Series in Statistics, 2nd edition, 1997.

Christopher K. Wikle, Andrew Zammit-Mangion, and Noel Cressie. *Spatio-Temporal Statistics with R*. Chapman and Hall/CRC, 1st edition, 2019. doi: https://doi.org/10.1201/9781351769723. URL https://spacetimewithr.org/.

Xiaoge Xin, Tongwen Wu, Jie Zhang, Junchen Yao, and Yongjie Fang. Comparison of cmip6 and cmip5 simulations of precipitation in china and the east asian summer monsoon. *International Journal of Climatology*, 40(15):6423–6440, 2020.

Farhad Yazdandoost, Sogol Moradian, Ardalan Izadi, and Amir Aghakouchak. Evaluation of cmip6 precipitation simulations across different climatic zones: Uncertainty and model intercomparison. *Atmospheric Research*, 250:105369, 2021.

Jiacheng Ye and Zhuo Wang. An origin of north american monsoon retreat biases in climate models. *Journal of Climate*, 36(14):4665–4676, 2023.

Sooin Yun, Xianyang Zhang, and Bo Li. Detection of local differences in spatial characteristics between two spatiotemporal random fields. *Journal of the American Statistical Association*, 117(537):291–306, 2022.

Yasin Zamani, Seyed Arman Hashemi Monfared, Mohsen Hamidianpour, et al. A comparison of cmip6 and cmip5 projections for precipitation to observational data: the case of northeastern iran. *Theoretical and Applied Climatology*, 142(3):1613–1623, 2020.

Jie Zhang, Tongwen Wu, Fang Zhang, Kalli Furtado, Xiaoge Xin, Xueli Shi, Jianglong Li, Min Chu, Li Zhang, Qianxia Liu, et al. Bcc-esm1 model datasets for the cmip6 aerosol chemistry model intercomparison project (aerchemmip). *Advances in Atmospheric Sciences*, 38:317–328, 2021.

Jin-Ting Zhang and Jianwei Chen. Statistical inferences for functional data. 2007.

Wenxia Zhang and Tianjun Zhou. Significant increases in extreme precipitation and the associations with global warming over the global land monsoon regions. *Journal of Climate*, 32(24):8465–8488, 2019.

Xianyang Zhang and Xiaofeng Shao. Two sample inference for the second-order property of temporally dependent functional data. 2015.

Shanshan Zhao, Wenping He, Tianyun Dong, Jie Zhou, Xiaoqiang Xie, Ying Mei, Shiquan Wan, and Yundi Jiang. Evaluation of the performance of cmip5 models to simulate land surface air temperature based on long-range correlation. *Frontiers in Environmental Science*, page 6, 2021.