ARTICLE IN PRESS

Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Spherical autoregressive models, with application to distributional and compositional time series☆

Changbo Zhu [a], Hans-Georg Müller [b,*]

[a] Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN 46556, USA
[b] Department of Statistics, University of California, Davis, Davis, CA 95616, USA

ARTICLE INFO

Article history:
Received 11 October 2021
Received in revised form 18 July 2022
Accepted 2 December 2022
Available online xxxx

Keywords:
Time series analysis
Fisher–Rao metric
Skew-symmetric operators
Rotation
Random objects

ABSTRACT

We introduce a new class of autoregressive models for spherical time series. The dimension of the spheres on which the observations of the time series are situated may be finite-dimensional or infinite-dimensional, where in the latter case we consider the Hilbert sphere. Spherical time series arise in various settings. We focus here on distributional and compositional time series. Applying a square root transformation to the densities of the observations of a distributional time series maps the distributional observations to the Hilbert sphere, equipped with the Fisher–Rao metric. Likewise, applying a square root transformation to the components of the observations of a compositional time series maps the compositional observations to a finite-dimensional sphere, equipped with the geodesic metric on spheres. The challenge in modeling such time series lies in the intrinsic non-linearity of spheres and Hilbert spheres, where conventional arithmetic operations such as addition or scalar multiplication are no longer available. To address this difficulty, we consider rotation operators to map observations on the sphere. Specifically, we introduce a class of skew-symmetric operators such that the associated exponential operators are rotation operators that for each given pair of points on the sphere map the first point of the pair to the second point of the pair. We exploit the fact that the space of skew-symmetric operators is Hilbertian to develop autoregressive modeling of geometric differences that correspond to rotations of spherical and distributional time series. Differences expressed in terms of rotations can be taken between the Fréchet mean and the observations or between consecutive observations of the time series. We derive theoretical properties of the ensuing autoregressive models and showcase these approaches with several motivating data. These include a time series of yearly observations of bivariate distributions of the minimum/maximum temperatures for a period of 120 days during each summer for the years 1990-2018 at Los Angeles (LAX) and John F. Kennedy (JFK) international airports. A second data application concerns a compositional time series with annual observations of compositions of energy sources for power generation in the U.S..

1. Introduction

Modern day data analysts increasingly encounter complex data types where data are no longer traditional vectors, and furthermore are not situated in a linear space such as a Hilbert space. Such non-Euclidean data may also be

☆ Research supported in part by NIH Echo UH3OD023313 and NSF DMS-2014626.
* Corresponding author.
E-mail address: hgmueller@ucdavis.edu (H.-G. Müller).

https://doi.org/10.1016/j.jeconom.2022.12.008

Please cite this article as: C. Zhu and H.-G. Müller, Spherical autoregressive models, with application to distributional and compositional time series. Journal of Econometrics (2023), https://doi.org/10.1016/j.jeconom.2022.12.008.

encountered in the form of a time series. At this point, the methodology available for the analysis of such data is quite limited. An exception are recent efforts to develop models for distributional time series in the context of the rapidly evolving field of distributional data analysis (DDA) (Petersen et al., 2021). A simple approach for distributional time series is to represent distributions by square integrable functions via the log quantile density transformation or a similar transformation (Petersen and Müller, 2016); a downside is that such transformations may lead to large metric distortions. The distributional time series is then transformed to a functional time series, which have been well investigated (Bosq, 2000). Geometric approaches that are based on constructing tangent bundles on the Wasserstein manifold have recently been shown to provide better predictions for autoregressive models (Chen et al., 2021; Zhang et al., 2022), while an autoregressive model that is intrinsic to the Wasserstein manifold can be based on a recently developed transport algebra (Zhu and Müller, 2021). It bears emphasizing that all these methods are limited to the case of distributional time series composed of one-dimensional distributions.

The reason that previous statistical modeling of distributional time series has been limited to the case of sequences of one-dimensional distribution lies in the challenges of characterizing optimal transport as well as the Wasserstein manifold and its parallel transport for the case of multivariate distributions. These challenges are formidable and there are also numerical difficulties. For both multivariate and one-dimensional distributions the Fisher–Rao metric provides an alternative to the popular Wasserstein metric that is easy to work with both numerically and theoretically, irrespective of the dimension of the distributions. This distributional metric is characterized by its invariance under diffeomorphisms and the ease of explicitly computing geodesics in the space of distributions with smooth densities equipped with this metric (Friedrich, 1991; Bauer et al., 2016). Of special interest for statistical applications is that the Fisher–Rao metric can be easily extended to multivariate distributions. Neither theoretical analysis nor numerical implementations face difficulties in the multivariate case and the geodesics in distribution space are always well-defined, irrespective of the dimension.

We focus here on time series data with observations that reside naturally or can be equivalently represented as points on a sphere $\mathcal{S} = \{x \in \mathcal{H} : \|x\|_{\mathcal{H}} = 1\}$, where $\mathcal{H}$ is a real separable Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and norm $\|x\|_{\mathcal{H}} := \sqrt{\langle x, x \rangle_{\mathcal{H}}}$. The sphere $\mathcal{S}$ can be finite-dimensional, in which case we denote it by $\mathcal{S}^d$ if $\mathcal{H} = \mathbb{R}^{d+1}$, or infinite-dimensional where $\mathcal{H} = L^2$ or any isomorphic space and in this case we refer to it as the Hilbert sphere and denote it by $\mathcal{S}^\infty$. Our focus on spherical time series is motivated by the convenience of incorporating different data types such as compositional data, directional data and distributional data into this novel framework.

In contrast to spherical time series, time series analysis for Euclidean vector data is a well-established field and both parametric and non-parametric methods have been fully developed (Fan and Yao, 2017, 2003). For functional or Hilbert-space valued time series autoregressive process models have also been well studied, starting with Bosq (2000). In contrast to these developments for Euclidean data, there is so far very little work on time series with random objects, i.e., random variables in general metric spaces (Müller, 2016). Even for the special case of spherical time series the literature is scarce. An interesting related approach is spherical regression that has been studied for the non-time series case when one has i.i.d. data with predictors and responses located in $\mathcal{S} = \mathcal{S}^d$ (Chang, 1986, 1989; Kim, 1998; Marzio et al., 2019), where the key ingredient is a rotation matrix in the set of orthogonal matrices $\mathrm{SO}(d+1)$ that rotates the predictor to the response. In addition, Downs (2003) and Rosenthal et al. (2014) introduced additional families of transformations and Shi et al. (2021) investigated settings where predictors and responses might have mismatches. However, all these methods are established under the i.i.d. regression setting and limited to the finite-dimensional case ($d < \infty$); furthermore, they allow for only one predictor, while for autoregressive modeling one needs to accommodate the joint action of predictors from multiple lags. The main challenge of modeling time series in non-linear spaces such as $\mathcal{S}$ is that conventional operations, including addition and subtraction, are not available, leading to a fundamental limitation for autoregressive modeling.

This paper provides three key innovations: First, to our knowledge, this is the very first paper that develops statistical modeling for spherical time series and specifically autoregressive models for such time series. Second, using the proposed approaches allows for regression models that feature multivariate distributions as predictors and responses, which was not possible with previous distributional regression models that stay within the space of distributions such as Wasserstein regression, due to the complex geometry of the Wasserstein space for multivariate distributions (Chen et al., 2021). While we develop such regression models here in the context of autoregressive modeling using the Fisher–Rao metric in distribution space, these models are also new for the independent case. Third, we provide the first modeling approach for spherical regression with multiple predictors for both independent and time series cases, as no spherical regression models currently exist that allow for multiple predictors. We emphasize that while we develop the second and third innovations in the context of autoregressive time series, the corresponding modeling and estimation for the independent case, where these are also new developments, is an immediate consequence.

The rest of the paper is organized as follows. Section 2 provides examples for spherical time series and perspectives as well as the central ideas of our modeling approach. In Section 3 we introduce rotation and log rotation operators in Hilbert spaces on which our model relies and present a key relationship between rotations and skew-symmetric operators. Methodology and theory are in Section 4, which also contains the main results. Estimation and prediction are studied in Section 5. We report results for data applications to distributional and compositional time series in Section 6, which is followed by a discussion section in Section 6.

2. Examples of spherical time series and basic approach

Compositional data take values in the simplex

$$\mathcal{C}^d = \left\{ \mathbf{z} = (z_1, z_2, \ldots, z_d)^T \in \mathbb{R}^d \;\middle|\; z_i \geq 0 \text{ for all } i = 1, 2, \ldots, d \text{ and } \sum_{i=1} z_i = 1 \right\},$$

reflecting that such data are non-negative proportions that sum to 1. By applying the point-wise square root ratio (psr) transformation $\mathrm{psr} : \mathcal{C}^d \to \mathcal{S}^{d-1}$, defined as

$$\mathrm{psr}(\mathbf{z}) = \left(\sqrt{z_1}, \sqrt{z_2}, \ldots, \sqrt{z_d}\right)^T, \tag{1}$$

$\mathcal{C}^d$ can be mapped to a subset of $\mathcal{S}^{d-1}$. This maps compositional time series to finite-dimensional spherical time series (Scealy and Welsh, 2011; Dai and Müller, 2018). Examples of compositional time series are common and include for example repeated election cycles when there are several parties and the compositions correspond to the vote shares of each party; or color preferences of car buyers that change from year to year, reflected in the percentage of cars sold in a specific color. In Section 5.2, we illustrate the proposed spherical autoregressive models (SAR) with compositional time series that correspond to the annually recorded proportions of electricity generated from different energy sources in the U.S., where energy sources include coal, natural gas or nuclear and renewable sources. The composition of energy sources has a major impact on the carbon dioxide (CO2) emissions that accrue from electricity generation over time.

Data that can be represented as locations on finite-dimensional spheres are ubiquitous and are not limited to compositional data but also include directional data such as wind directions. For example, the study of ocean surface wind over time is important in determining the spread of aerial organisms (Felicísimo et al., 2008). Sequences of hourly or daily recorded wind directions are naturally represented as a spherical time series with observations in $\mathcal{S}^1$, the unit circle. Another application of $\mathcal{S}^d$-valued time series are vector time series, where the vector observations can be expressed in polar coordinates and then form a spherical time series and a scalar time series, where the latter corresponds to the length of the vector at time t. In some cases the length of the vector may not be relevant if one is primarily interested in the association between the vector components as reflected by the direction of the vector. Then only the sequence of directions of the vector components matters and if the original vector data have dimension $d + 1$, the directions are represented on $\mathcal{S}^d$, giving rise to a spherical time series.

For any density function $f : \mathbb{R}^D \to \mathbb{R}$, where the dimension of the domain D is a positive integer, f satisfies $f \geq 0$ and $\int_{\mathbb{R}^D} f(x)dx = 1$, and we define the functional point-wise square root transformation (fpsr) as

$$\mathrm{fpsr}(f) = g, \text{ where } g(\mathbf{z}) = \sqrt{f(\mathbf{z})} \text{ for all } \mathbf{z} \in \mathbb{R}^D.$$

Using fpsr, distributional data correspond to the elements of a segment of the Hilbert sphere $\mathcal{S}^\infty$ equipped with the Fisher–Rao metric (Dai, 2022) and distributional time series then are represented as $\mathcal{S}^\infty$-valued time series. An example are two-dimensional distributions of daily maximum and minimum temperatures recorded for 24 h over the summer months at airports in the U.S. Considering these two-dimensional distributions over successive years then forms a time series with $\mathcal{S}^\infty$-valued observations. These time series are of interest for assessing the effects of climate change and the risks and costs associated with rising temperatures.

To overcome the challenge of non-linearity for the case of spherical time series, we utilize the geometric structure of spheres $\mathcal{S}$, where the ambient vector space into which the sphere is embedded is $\mathbb{R}^d$ or L^2. The geodesic distance on $\mathcal{S}$ is defined as $d(x_1, x_2) = \arccos(\langle x_1, x_2 \rangle)$ for any $x_1, x_2 \in \mathcal{S}$, where the inner product is defined in the ambient vector space. Geodesics are locally length-minimizing paths between points that are well-defined in geodesic metric spaces, where the length of a geodesic path between two points coincides with the distance of the points (Burago et al., 2001). The geodesics of spheres correspond to great circles. The key idea for the modeling of spherical autoregressive (SAR) time series is that the geodesic between two points $x_1, x_2 \in \mathcal{S}$ can be written as $\gamma(a) = \exp(aL)$, where $a \in [0, 1]$ and $L : \mathcal{H} \to \mathcal{H}$ is a skew-symmetric operator. We then relate the spherical difference between x_1 and x_2 to the linear operator L. This makes it possible to model the differenced times series in the linear space of skew-symmetric operators.

We study two versions of autoregressive models for spherical time series. In the basic SAR model the autoregressive model is based on the spherical equivalent of differences between the observations and the overall Fréchet mean, while DSAR is a variant based on the spherical differences between consecutive observations. These models can be applied for autoregressive modeling on any type of sphere $\mathcal{S}$ and their implementation is computationally efficient.

3. Rotations and skew-symmetric operators

Let $\mathcal{H}$ be a real separable Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and norm $\|x\|_{\mathcal{H}} := \sqrt{\langle x, x \rangle_{\mathcal{H}}}$. The Hilbert sphere $\mathcal{S}$ is a subset of $\mathcal{H}$ whose elements have norm 1, i.e., $\mathcal{S} = \{x \in \mathcal{H} : \|x\|_{\mathcal{H}} = 1\}$. Given a set of points $\{x_1, x_2, \ldots, x_m\} \subset \mathcal{S}$, let $\mathrm{span}\{x_1, x_2, \ldots, x_m\} = \{a_1x_1 + a_2x_2 + \cdots + a_mx_m : a_1, a_2, \ldots, a_m \in \mathbb{R}\} \subset \mathcal{H}$ denote the m-dimensional subspace of $\mathcal{H}$ spanned by $x_1, x_2, \ldots, x_m$. The set of bounded linear operators on $\mathcal{H}$ is denoted as $\mathscr{B}(\mathcal{H})$ and an operator $Q \in \mathscr{B}(\mathcal{H})$ is skew-symmetric if

$$\langle Qx, y \rangle + \langle x, Qy \rangle = 0 \text{ for all } x, y \in \mathcal{H}.$$

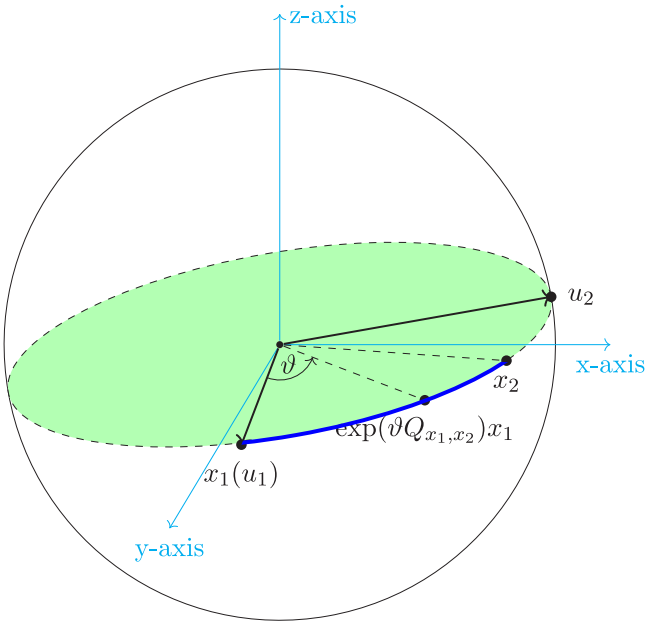

Fig. 1. Illustration of the rotation operator $\exp(\vartheta Q_{x_1,x_2})$ when $\mathcal{H} = \mathbb{R}^3$. The green plane is the two-dimensional subspace spanned by u_1 and u_2. By construction, u_1, u_2 are orthogonal and the angle between them is $\pi/2$. Here $\exp(\vartheta Q_{x_1,x_2})x_1$ is the location of the image of the rotation operator $\exp(\vartheta Q_{x_1,x_2})$ applied at x_1 and ϑ is the angle between x_1 and $\exp(\vartheta Q_{x_1,x_2})x_1$. The blue line is the geodesic between x_1 and x_2 that is traced by the path $\gamma(a) := \exp(a\theta Q_{x_1,x_2})x_1$, where $a \in [0, 1]$ and $\theta = \arccos(\langle x_1, x_2\rangle)$. It can be easily seen that $\gamma(0) = x_1$ and $\gamma(1) = x_2$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

For any bounded linear operator $L : \mathcal{H} \to \mathcal{H}$, consider its exponential $\exp(L) := \sum_{l=0}^{\infty} L^l/l!$. An orthogonal operator $O \in \mathscr{B}(\mathcal{H})$ is a rotation operator if and only if there exists a skew-symmetric operator Q such that $O = \exp(Q)$ (Martin, 1932).

For each skew-symmetric operator Q there is a unique rotation operator $\exp(Q)$. Let $\mathscr{R}(\mathcal{H})$ and $\mathscr{S}(\mathcal{H})$ be the set of rotation operators and skew-symmetric operators, respectively; by definition $\mathscr{R}(\mathcal{H}) = \exp(\mathscr{S}(\mathcal{H}))$. If $\{e_1, e_2, \ldots\}$ is an orthonormal basis of $\mathcal{H}$, then $\mathscr{S}(\mathcal{H})$ admits the orthonormal basis

$$\mathscr{S}(\mathcal{H}) = \text{span}\left\{e_i \otimes e_j - e_j \otimes e_i : i, j = 1, 2, \ldots\right\} \subset \mathcal{H} \otimes \mathcal{H},$$

where $\mathcal{H} \otimes \mathcal{H}$ is the tensor product of the Hilbert space $\mathcal{H}$ with itself and is also a Hilbert space with inner product

$$\langle x_1 \otimes y_1, x_2 \otimes y_2\rangle_{\mathcal{H}\otimes\mathcal{H}} = \langle x_1, x_2\rangle_{\mathcal{H}}\langle y_1, y_2\rangle_{\mathcal{H}},$$

with $x_1, x_2, y_1, y_2 \in \mathcal{H}$; the inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}\otimes\mathcal{H}}$ can be extended to any element in $\mathcal{H} \otimes \mathcal{H}$ by linearity. Observing that $\mathscr{S}(\mathcal{H})$ is a closed subspace of $\mathcal{H} \otimes \mathcal{H}$ with respect to $\langle\cdot,\cdot\rangle_{\mathcal{H}\otimes\mathcal{H}}$, $\mathscr{S}(\mathcal{H})$ is seen to be a complete separable Hilbert space.

Given two points $x_1, x_2 \in \mathcal{H}$ such that $x_1 \neq x_2$ and $x_1 \neq -x_2$, the proposed methodology relies on rotation operators that provide a rotation on $\mathcal{S}$ within the two dimensional subspace $\text{span}\{x_1, x_2\}$.

Theorem 1. *Set $u_1 = x_1$ and $u_2 = (x_2 - \langle x_2, u_1\rangle u_1)/\|x_2 - \langle x_2, u_1\rangle u_1\|_{\mathcal{H}}$. Let I be the identity operator and $Q_{x_1,x_2} := u_1 \otimes u_2 - u_2 \otimes u_1 \in \mathscr{S}(\mathcal{H})$. Then, given an angle $\vartheta \in [0, 2\pi]$,*

$$\exp(\vartheta Q_{x_1,x_2}) = I + sin(\vartheta)Q_{x_1,x_2} + (1 - cos(\vartheta))Q_{x_1,x_2}^2 \tag{2}$$

is a rotation operator that rotates counterclockwise within $\text{span}\{u_1, u_2\}$ by ϑ, i.e.,

- *For any $y_1, y_2 \in \mathcal{H}$, $\langle\exp(\vartheta Q_{x_1,x_2})y_1, \exp(\vartheta Q_{x_1,x_2})y_2\rangle_{\mathcal{H}} = \langle y_1, y_2\rangle_{\mathcal{H}}$.*
- *For any $x \in \text{span}\{u_1, u_2\} \cap \mathcal{S}$, $\arccos\langle\exp(\vartheta Q_{x_1,x_2})x, x\rangle_{\mathcal{H}} = \vartheta$.*
- *For any $y \in \mathcal{H}$ perpendicular to $\text{span}\{u_1, u_2\}$, i.e., $\langle y, u_1\rangle_{\mathcal{H}} = 0$ and $\langle y, u_2\rangle_{\mathcal{H}} = 0$, it holds that $\exp(\vartheta Q_{x_1,x_2})y = y$.*

For $\mathcal{H} = \mathbb{R}^3$, Fig. 1 provides an illustration of the rotation operator $\exp(\vartheta Q_{x_1,x_2})$. We note that (2) reduces to the Rodrigues rotation formula in this special case. For a rotation operator $\exp(\vartheta L)$ in higher dimensional Hilbert spaces such as $\mathbb{R}^d$ with $d > 3$, where L is an arbitrary skew-symmetric operator, the equality $\exp(\vartheta L) = I + \sin(\vartheta)L + (1 - \cos(\vartheta))L^2$ will not hold in general. That (2) is satisfied for any separable space $\mathcal{H}$ is due to the fact that $\exp(\vartheta Q_{x_1,x_2})$ is a special rotation that only rotates within the two-dimensional subspace $\text{span}\{x_1, x_2\}$.

4. Spherical autoregressive models

Based on the rotation operator introduced earlier, the geodesic $\gamma : [0, 1] \to \mathcal{S}$ between two points $x_1, x_2 \in \mathcal{S}$ can be traced by rotating x_1 to x_2 within the two dimensional subspace spanned by $\{x_1, x_2\}$ around the origin, i.e.,

$$\gamma(a) = \exp(a\theta Q_{x_1,x_2})x_1. \tag{3}$$

where $a \in [0, 1]$, $\theta = \arccos(\langle x_1, x_2\rangle)$ is the angle between x_1 and x_2 and Q_{x_1,x_2} is as defined in Theorem 1; see Fig. 1 for a demonstration of $\gamma(a)$ when $\mathcal{H} = \mathbb{R}^3$. We then utilize geodesics on $\mathcal{S}$ to arrive at a notion of difference between points on $\mathcal{S}$.

Starting with the Euclidean space $\mathbb{R}^d$ and considering two elements $w_1, w_2 \in \mathbb{R}^d$, the difference $V = w_2 - w_1$ can be interpreted as an optimal (in the sense of minimizing Euclidean distance) transport map that moves w_1 to w_2 and the connecting geodesic in Euclidean space is represented as the straight line $r(a) = w_1 + aV$ where $a \in [0, 1]$, with corresponding transport on the sphere given by the geodesic (3), which is $\gamma(a) = \exp(a\theta Q_{x_1,x_2})$; see Zhu and Müller (2021) for a similar extension of this idea to the Wasserstein space. Therefore, recalling that $\theta = \arccos(\langle x_1, x_2\rangle)$ is the angle between x_1 and x_2, $Q_{x_1,x_2} := u_1 \otimes u_2 - u_2 \otimes u_1 \in \mathscr{S}(\mathcal{H})$ is the rotation operator that rotates x_1 to x_2 and $u_1 = x_1$, $u_2 = (x_2 - \langle x_2, u_1\rangle u_1)/\|x_2 - \langle x_2, u_1\rangle u_1\|_{\mathcal{H}}$, it makes sense to define a spherical difference between points x_1 and x_2 on $\mathcal{S}$ as

$$x_2 \ominus x_1 := \theta Q_{x_1,x_2}. \tag{4}$$

Given a sequence of data points $x_1, x_2, \ldots, x_n \in \mathcal{S}$ with the same Fréchet mean μ_x, i.e., $\mu_x := \operatorname{argmin}_{z\in\mathcal{S}} \mathbb{E}[d^2_{\mathcal{S}}(z, x_t)]$ for all $t = 1, 2, \ldots, n$, we then construct a new series by taking differences between the x_t and the Fréchet mean μ_x,

$$\{R_t := x_t \ominus \mu_x : t = 1, 2, \ldots, n\} \subset \mathscr{S}(\mathcal{H}).$$

Assuming that $\{R_t\}$ is a stationary sequence (Bosq, 2000), we propose the following spherical autoregressive (SAR) model of order p,

$$R_t - \mu_R = \alpha_1(R_{t-1} - \mu_R) + \cdots + \alpha_p(R_{t-p} - \mu_R) + \varepsilon_t \text{ where } R_t = x_t \ominus \mu_x, \tag{5}$$

where $\alpha_1, \ldots, \alpha_p \in \mathbb{R}$, $\mu_R = E[R_t]$ and $\{\varepsilon_t\} \subset \mathscr{S}(\mathcal{H})$ are i.i.d random innovations with mean 0.

To elucidate the connection of this model with the previously studied spherical regression (Chang, 1986, 1989; Kim, 1998; Marzio et al., 2019), which has not yet been extended to a time series framework and admits only one predictor, consider a regression setting with x_t as single predictor and y_t as response. In the above difference notation (4), this previously studied spherical regression model can be written as $y_t \ominus x_t = R_0 + \varepsilon_t$. In the Euclidean space $\mathbb{R}^d$ this corresponds to an intercept only regression model $z_t - w_t = \beta_0 + \epsilon_t$, where $z_t \in \mathbb{R}^d$ is the response, $w_t \in \mathbb{R}^d$ is the predictor, $\beta_0 \in \mathbb{R}^d$ is the intercept and $\{\epsilon_t\} \subset \mathbb{R}^d$ are i.i.d. errors. By taking expectation on both sides, we observe that $E[z_t] - E[w_t] = \beta_0$. In some sense, this corresponds to a special case of Model (5) where $p = 1$ and the single "slope" is $\alpha_1 = 1$ as then one obtains $y_t \ominus \mu_y = x_t \ominus \mu_x + \varepsilon_t$ and its Euclidean counterpart $z_t - E[z_t] = w_t - E[w_t] + \epsilon_t$, which is equivalent to $z_t - w_t = \beta_0 + \epsilon_t$.

As alternative to the SAR model (5) we also consider a second model that is based on the spherical differences of consecutive observations. This difference based spherical autoregressive model (DSAR) is given by

$$R_t - \mu_R = \alpha_1(R_{t-1} - \mu_R) + \cdots + \alpha_p(R_{t-p} - \mu_R) + \varepsilon_t, \text{ where } R_t = x_{t+1} \ominus x_t, \tag{6}$$

where as before, $\alpha_1, \ldots, \alpha_p \in \mathbb{R}$, $\mu_R = E[R_t]$ and $\{\varepsilon_t\} \subset \mathscr{S}(\mathcal{H})$ are i.i.d random innovations with mean 0.

Differencing is an inherent feature of DSAR models and is a common technique to reduce trend and seasonality for nonstationary time series in Euclidean space. It may also be useful for some spherical time series. For example, the US energy mix compositional time series, which we will discuss further in Section 5.2, shows a trend over the years, as more clean energy is generated each year and coal/petroleum fuels are increasingly phased out.

We note that both model (5) and model (6) have scalar coefficients. When $\mathcal{H}$ is an infinite-dimensional space, we can extend the two models by considering bounded linear operators as coefficients, i.e.,

$$R_t - \mu_R = L_1(R_{t-1} - \mu_R) + \cdots + L_p(R_{t-p} - \mu_R) + \varepsilon_t,$$

where $L_1, \ldots, L_p \in \mathscr{B}(\mathcal{H})$. Regarding the existence of stationary solutions of the proposed SAR model, the following result is a consequence of Theorem 3.3 of Zhang et al. (2022).

Theorem 2. *Assuming that $\{R_t : t \in \mathbb{N}\}$ is stationary, $E\langle\varepsilon_t, \varepsilon_t\rangle_{\mathcal{H}\otimes\mathcal{H}} < \infty$ and the roots of $\phi(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p$ are outside the unit circle, then*

$$R_t - \mu_R = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

is a unique stationary solution of

$$R_t - \mu_R = \alpha_1(R_{t-1} - \mu_R) + \cdots + \alpha_p(R_{t-p} - \mu_R) + \varepsilon_t, \ t \in \mathbb{N},$$

where $\{\psi_t\}$ is absolutely summable and determined by $1/\phi(z) = \sum_{i=0}^{\infty} \psi_i z^i$.

The above theorem uses the stationarity of $\{R_t\}$. It follows from the definition of $\{R_t\}$, i.e., $R_t = x_t \ominus \mu_x$ in SAR and $R_t = x_{t+1} \ominus x_t$ in DSAR, that the stationarity of x_t will imply the stationarity of $\{R_t\}$.

5. Estimation and prediction

5.1. Estimation

We use Yule–Walker type estimators for the estimation of the coefficients $\alpha_1, \alpha_2, \ldots, \alpha_p$ of the SAR and DSAR models. Setting

$$\lambda_k = E[\langle R_1 - \mu_R, R_{k+1} - \mu_R\rangle_{\mathcal{H}\otimes\mathcal{H}}],$$

it is straightforward to check that the model parameters satisfy

$$\begin{pmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_p\end{pmatrix} = \begin{pmatrix}\lambda_0 & \lambda_1 & \cdots & \lambda_{p-1}\\ \lambda_1 & \lambda_0 & \cdots & \lambda_{p-2}\\ \vdots & \vdots & & \vdots\\ \lambda_{p-1} & \lambda_{p-2} & \cdots & \lambda_0\end{pmatrix}\begin{pmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_p\end{pmatrix}.$$

Replacing λ_k by sample estimates

$$\widehat{\lambda}_k = \frac{1}{n-k}\sum_{t=1}^{n-k}\langle R_t - \widehat{\mu}_R, R_{t+k} - \widehat{\mu}_R\rangle_{\mathcal{H}\otimes\mathcal{H}}, \quad \widehat{\mu}_R = \frac{1}{n}\sum_{t=1}^{n} R_t \tag{7}$$

then suggests the following estimates $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_p$ for the model parameters $\alpha_1, \ldots, \alpha_p$,

$$\begin{pmatrix}\widehat{\alpha}_1\\ \widehat{\alpha}_2\\ \vdots\\ \widehat{\alpha}_p\end{pmatrix} = \begin{pmatrix}\widehat{\lambda}_0 & \widehat{\lambda}_1 & \cdots & \widehat{\lambda}_{p-1}\\ \widehat{\lambda}_1 & \widehat{\lambda}_0 & \cdots & \widehat{\lambda}_{p-2}\\ \vdots & \vdots & & \vdots\\ \widehat{\lambda}_{p-1} & \widehat{\lambda}_{p-2} & \cdots & \widehat{\lambda}_0\end{pmatrix}^{-1}\begin{pmatrix}\widehat{\lambda}_1\\ \widehat{\lambda}_2\\ \vdots\\ \widehat{\lambda}_p\end{pmatrix}. \tag{8}$$

Writing $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_p)^T$ and $\widehat{\boldsymbol{\lambda}} = (\widehat{\lambda}_0, \widehat{\lambda}_1, \ldots, \widehat{\lambda}_p)^T$, we next establish asymptotic normality for $\widehat{\boldsymbol{\lambda}}$.

Theorem 3. *Under the assumptions of Theorem 2, it holds that*

$$\sqrt{n}(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \to^d N(0, V), \qquad V = \left(\sum_{h=-\infty}^{\infty}\Gamma_{u,v}^h\right)_{u,v=0,1,\ldots,p},$$

where, setting $\kappa(u) = \sum_{i=-\infty}^{\infty}\psi_i\psi_{i+u}$,

$$\Gamma_{u,v}^h = \left(E[\langle\varepsilon_1,\varepsilon_1\rangle^2] - (E[\langle\varepsilon_1,\varepsilon_1\rangle])^2 - 2E[\langle\varepsilon_1,\varepsilon_2\rangle^2]\right)\sum_{i=-\infty}^{\infty}\psi_i\psi_{i+u}\psi_{i+h}\psi_{i+h+v}$$
$$+(E[\langle\varepsilon_1,\varepsilon_1\rangle])^2\kappa(u)\kappa(v) + E[\langle\varepsilon_1,\varepsilon_2\rangle^2]\left(\kappa(h)\kappa(h+v-u) + \kappa(h+v)\kappa(h-u)\right).$$

For the case where $\{\varepsilon_t\}$ are i.i.d random innovations in $\mathbb{R}$, the $\Gamma_{u,v}^h$ are identical to those in Bartlett's formula. To show the convergence of $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_p$, we set

$$\Lambda = \begin{pmatrix}\lambda_0 & \lambda_1 & \cdots & \lambda_{p-1}\\ \lambda_1 & \lambda_0 & \cdots & \lambda_{p-2}\\ \vdots & \vdots & & \vdots\\ \lambda_{p-1} & \lambda_{p-2} & \cdots & \lambda_0\end{pmatrix} \text{ and } \widehat{\Lambda} = \begin{pmatrix}\widehat{\lambda}_0 & \widehat{\lambda}_1 & \cdots & \widehat{\lambda}_{p-1}\\ \widehat{\lambda}_1 & \widehat{\lambda}_0 & \cdots & \widehat{\lambda}_{p-2}\\ \vdots & \vdots & & \vdots\\ \widehat{\lambda}_{p-1} & \widehat{\lambda}_{p-2} & \cdots & \widehat{\lambda}_0\end{pmatrix}.$$

If $\det(\Lambda) \neq 0$, it follows from the continuous mapping theorem that $\widehat{\Lambda}^{-1} \to^p \Lambda^{-1}$ and thus by Theorem 3,

Corollary 1. *Under the assumptions of Theorem 2, if* $\det(\Lambda) \neq 0$,

$$\sqrt{n}\left(\begin{pmatrix}\widehat{\alpha}_1\\ \vdots\\ \widehat{\alpha}_p\end{pmatrix} - \begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_p\end{pmatrix}\right) \to^d N\left(0, \Lambda\widetilde{V}(\Lambda^T)^{-1}\right), \text{ where } \widetilde{V} = \left(\sum_{h=-\infty}^{\infty}\Gamma_{u,v}^h\right)_{u,v=1,\ldots,p}.$$

We note that in applications involving distributional time series the distributions and specifically the density functions f_t are usually not directly observed and must be inferred from available samples of size N_t, $\{z_{i,t} \in \mathbb{R}^D : i =$

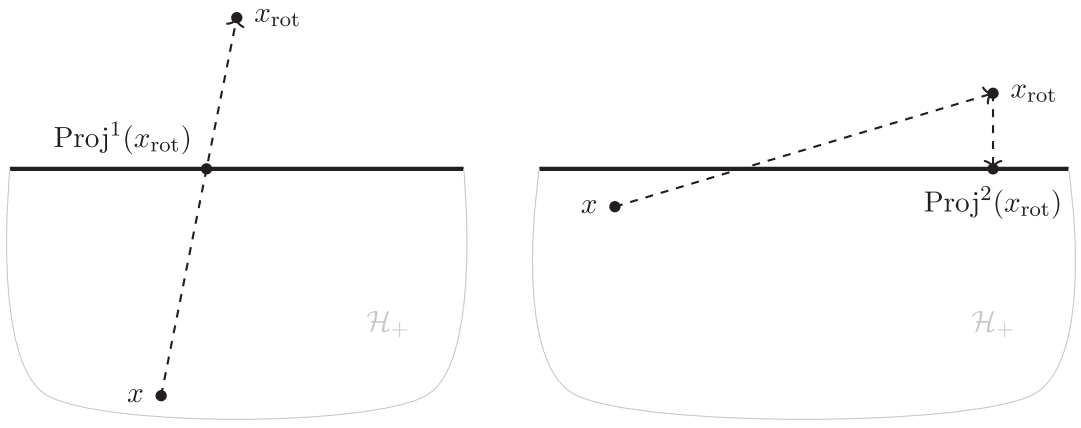

Fig. 2. Illustration of the two projection operators Proj^1 and Proj^2.

$1, 2, \ldots, N_t\} \sim^{i.i.d} f_t$ that they generate, separately for each t. The random mechanism that generates the samples is assumed to be independent from the mechanism that generates the random distributions.

To assess the impact of this preliminary estimation step requires additional assumptions as follows: All densities f_t have the same compact domain $A \subset \mathbb{R}^D$ and are continuously differentiable on their support; and there is a sequence $N \to \infty$ such that $N_t \geq N$ for all t; there exists a constant M such that for all t, $\sup_{a\in A} |f_t(a)|$, $\sup_{a\in A} 1/|f_t(a)|$, $\sup_{a\in A} \|f_t'(a)\|$ are all bounded by M, where $\|f_t'(a)\|$ is the norm of the gradient vector. Extending the arguments and construction in Petersen and Müller (2016) to the case of multivariate distributions then leads to density estimators $\hat{f}_t$ that satisfy

$$\sup_t P\left(\sup_{a\in A}\left|\hat{f}_t(a) - f_t(a)\right| > c_1 N^{-c_2}\right) \to 0$$

for constants $c_1, c_2 > 0$, where c_2 depends on the dimension of the distributions and decreases when the dimension increases. Making the additional assumption that the sample size $N = N(n)$ available for the estimation of each density f_t is such that $N^{-c_2} = o_p(n^{-1/2})$, one can then show that when substituting $\hat{f}_t$ for f_t in R_t in models (5) and (6), Theorem 3 and Corollary 1 still hold when using estimated instead of true densities.

5.2. Prediction

With estimates $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_p$ (7), (8) based on a data sequence $\{R_1, \ldots, R_n\}$ in hand, the prediction for the skew-symmetric operator at time $n+1$ is

$$\widehat{R}_{n+1} = \widehat{\mu}_R + \widehat{\alpha}_1(R_n - \widehat{\mu}_R) + \cdots + \widehat{\alpha}_p(R_{n-p+1} - \widehat{\mu}_R),$$

with a slight abuse of notation, as in model DSAR, the sequence of observations available for the prediction is of length $n+1$, i.e., $\{x_i : i = 1, 2, \ldots, n+1\}$, whereas in model SAR it is of length n. Once $\widehat{R}_{n+1}$ has been obtained, the prediction of the next observation in the original time series is $\widehat{x}_{n+1} := \exp(\widehat{R}_{n+1})\mu_x$ for SAR and $\widehat{x}_{n+2} := \exp(\widehat{R}_{n+1})x_{n+1}$ for DSAR.

For a distributional time series with D-dimensional distributions (or density functions), we set $\mathcal{H} = \{f : \mathbb{R}^D \to \mathbb{R} \mid \int_{\mathbb{R}^D} f^2(a)da < \infty\}$ with inner product $\langle f, g\rangle_{\mathcal{H}} = \int_{\mathbb{R}^D} f(a)g(a)da$ and require the predictions to be constrained in the positive orthant $\mathcal{H}_+ := \{f \in \mathcal{H} : f(a) \geq 0 \text{ for all } a \in \mathbb{R}^D\}$. Similarly, for compositional time series one has $\mathcal{H} = \mathbb{R}^d$ and the prediction is constrained to lie in $\mathcal{H}_+ = \{\mathbf{z} = (z_1, \ldots, z_d)^T \in \mathbb{R}^d : z_i \geq 0 \text{ for all } i\}$. Writing $x_{\rm rot} = \exp(Q)x$ for the rotation $\exp(Q)$ of $x \in \mathcal{H}_+$, we use projection operators to enforce the constraint; see Chen et al. (2021) and Pegoraro and Beraha (2022) for related projections in Wasserstein space. For the projection operators we consider two options.

A first option is to use a projection operator Proj^1 to rotate x_{rot} back to the boundary of $\mathcal{H}_+$, i.e., $\mathrm{Proj}^1(x_{\rm rot}) := \exp(c_1 Q)x$, where $c_1 = \sup\{c : c \in [0, 1] \text{ and } \exp(cQ)x \in \mathcal{H}_+\}$. A second option is to employ an operator Proj^2 to project $x_{\rm rot}$ to the nearest point in $\mathcal{H}_+$, i.e., $\mathrm{Proj}^2(x_{\rm rot}) = \mathrm{argmin}_{y\in\mathcal{H}_+}\langle x_{\rm rot} - y, x_{\rm rot} - y\rangle$; see Fig. 2 for a schematic illustration. We note that Proj^1 may be more useful for SAR, as all the predictions are based on rotations from the Fréchet mean, which may be more likely to stay away from the boundary of $\mathcal{H}_+$ under stationarity assumptions. Applying Proj^1 when constructing predictions of SAR leads to constrained predictions that are closer to the Fréchet mean than those obtained with Proj^2. On the other hand, Proj^1 may be less useful for DSAR, as one may obtain $\widehat{x}_{n+2} := \exp(\widehat{R}_{n+1})x_{n+1} \approx x_{n+1}$. Therefore Proj^2 appears to be more suitable for DSAR. In the following, we use Proj^1 for SAR and Proj^2 for DSAR.

6. Applications

6.1. Temperature data

Global warming is expected to lead to more heat waves in the summer. It is then of interest to study and model the time series of the bivariate distributions of daily minimum and maximum temperature. Extreme temperatures are associated with increased health and economic risks. The analysis reported here was inspired by Bhatia and Katz (2021). The temperature data we used have been recorded at airport weather stations in the U.S. over the years and are available at https://www.ncdc.noaa.gov/cdo-web/search?datasetid=GHCND.

On the ith day of year $t = 1990, \ldots, 2019$, we observe two temperatures $(z_{t,i}, w_{t,i})$, where $z_{t,i}$, $w_{t,i}$ are the minimum and maximum temperature over each 24 h period, respectively. We assume that the distribution of $(z_{t,i}, w_{t,i})$ over the summer months in year t has a two-dimensional density f_t such that

$$\{(z_{t,i}, w_{t,i}) : i = 1, 2, \ldots, N\} \sim^{i.i.d} f_t, \tag{9}$$

where $N = 122$, as the summer period is from June 1 to September 30. Here, $x_t = f_t$ and $\mathcal{H} = L^2(\mathcal{T})$, where $\mathcal{T} \subset \mathbb{R}^2$ is the temperature range.

In a preprocessing step we obtained estimates of the bivariate density functions f_t based on samples (9). A quick and fast smoother that adjusts for boundaries is histogram smoothing, which we implemented with histogram bins of size 50×50. We then applied the R package "fdapace" (Gajardo et al., 2021) for the smoothing step, where the bandwidths were chosen as $(\max_i z_{t,i} - \min_i z_{t,i})/5$ and $(\max_i w_{t,i} - \min_i w_{t,i})/5$, then adjusted the results so that the estimates are bona fide densities. We thus obtained 30 bivariate density functions for the years from 1990 to 2019, some of which are shown as contour plots in the top six panels of Figs. 3 and 4 for Los Angeles International Airport (LAX) and John F. Kennedy International Airport (JFK) respectively. We used the observed density for 2019 to illustrate the predictions obtained with SAR and DSAR using only the data before 2019 to construct the prediction. The predicted densities are shown as contour plots at the bottom of Figs. 3 and 4, where we chose the order $p = 5$ for both SAR and DSAR. We conclude from both the contour plots and the Fisher–Rao distances that SAR works better than DSAR for this prediction. In addition, we plotted the FR distances between the observed and the fitted densities across time in Fig. 5. There is no obvious trend, indicating a basic level of stationarity. Interestingly, the temperature sequence for LAX contains an outlier for 2012, a year with the highest temperature on record (113 °F) since 1921.

6.2. Energy data

Data on the sources of energy expressed as fractions or percentages for electricity generation across the entire U.S. are available at https://www.eia.gov/electricity/data/state/ and constitute a compositional time series. For our analysis we consider three energy sources: (i) Coal or Petroleum; (ii) Natural gas; (iii) Nuclear and Renewables. Sources (i) are known to produce the highest amounts of CO_2 and health damaging air pollutants per Watt generated, while sources (ii) are cleaner but still produce sizable amounts of CO_2. Sources (iii) do not directly produce damaging gases but may have some residual risks such as nuclear energy production. Here we consider the compositional time series consisting of the annual proportions of energy generated from sources (i)-(iii), which thus has three components.

The data are available for the years $t = 2005, 2006, \ldots, 2019$ and we denote the resulting time series by (U_t, V_t, W_t), where $U_t, V_t, W_t \geq 0$ and $U_t + V_t + W_t = 1$ for all t. We then obtain the spherical time series $x_t = (\sqrt{U_t}, \sqrt{V_t}, \sqrt{W_t}) \in \mathcal{S}^2$. For this example, we have $\mathcal{H} = \mathbb{R}^3$. The data $\{x_t\}_{t=2005}^{2018}$ are used as training set to fit SAR and DSAR models and we aim to predict the proportions of the energy sources for the year 2019. The observed compositions from 2005 to 2018 and the observed, fitted and predicted compositions for 2019 are shown in Fig. 6 and illustrated with two types of graphical representations for compositional data. A ternary plot is in the top panel and spherical plot in the bottom panel, where for the latter we plotted the longitude and latitude of each point $x_t \in \mathcal{S}^2$.

Both plots show a strong trend over the years and the ternary plot indicates that the proportion of energy generated from source (iii) is continuously increasing each year. Correspondingly, the proportion of energy from coal or petroleum is continuously decreasing. The trend indicates some degree of non-stationarity of x_t, while no trend seems to be present when considering the annual increments that correspond to the spherical rotations from one year to the next. It thus appears that the differences $\{x_{t+1} \ominus x_t\}$ are sufficiently stationary. Consequently, we applied model DSAR. Fig. 6 indicates that DSAR not only fits the observed data quite well but also produces a reasonable prediction for the energy mix in the year 2019. For comparison, we also applied the vector autoregressive model (VAR) for these data. The order of VAR was selected using an AIC criterion as implemented in Pfaff (2008). This approach ignores the spherical nature of the data. The Fisher–Rao distance between predicted and observed with the leave-out method is 0.046, while the proposed DSAR method, also implemented with the leave-out method, has a Fisher–Rao distance of 0.022. Further details about leave-out cross-validation for order selection are provided in the next section.

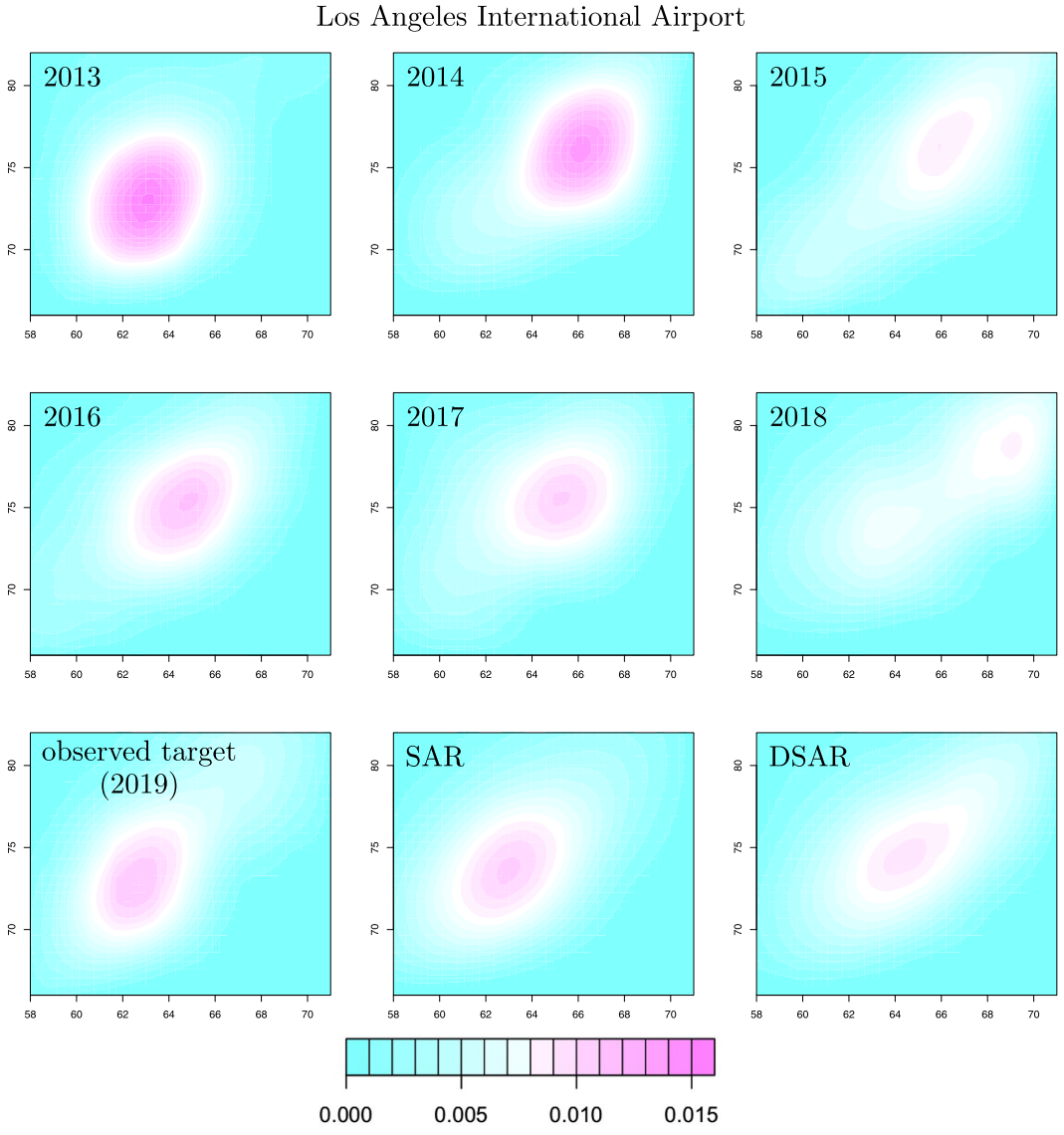

Fig. 3. Contour plots of observed and predicted two-dimensional density functions for the distributional time series of temperatures as recorded at LAX. The top six panels show the observed density functions in the training set. The bottom left panels show the observed distribution for 2019 (left); the predicted density using SAR (middle), with Fisher–Rao distance between predicted and observed of 0.197; and the predicted density using DSAR, with Fisher–Rao distance 0.236.

7. Simulation results

We demonstrate that ignoring the spherical structure of the data may lead to inferior performance by using simulated data. For a fair comparison, the data example used in this section is generic and not generated from our model. Let $\{\varepsilon_t\}$ be i.i.d. 6-dimensional Gaussian with mean 0 and covariance matrix Σ_ε. We generate a sequence of points $\{x_i\}$ on $\mathcal{S}^5$ as follows,

$$\begin{aligned} y_i &= \mu + \alpha_1 x_{i-1} + \alpha_2 x_{i-2} + \alpha_3 x_{i-3} + \varepsilon_i, \\ x_i &= y_i / \|y_i\|_2, \end{aligned} \tag{10}$$

where we consider two models for $\{\varepsilon_i\}$: (i) $\mu = (1, 0, 0, 0, 0, 0)^T$ and $\Sigma_\varepsilon = (\sigma_{ij})_{i,j=1}^6$, where $\sigma_{ij} = 0.25$ if $i = j$ and otherwise $\sigma_{ij} = 0$; (ii) $\mu = (0, 0, 0, 0, 0, 0)^T$ and $\Sigma_\varepsilon = (\sigma_{ij})_{i,j=1}^6$, where $\sigma_{ij} = \exp(-|i-j|)/4$.

To apply SAR or DSAR, we select the order p of the corresponding model using a cross-validation type criterion. For instance, given $\{x_1, \ldots, x_n\}$, we can use a portion of the data as test set and adopt the rolling window approach. More

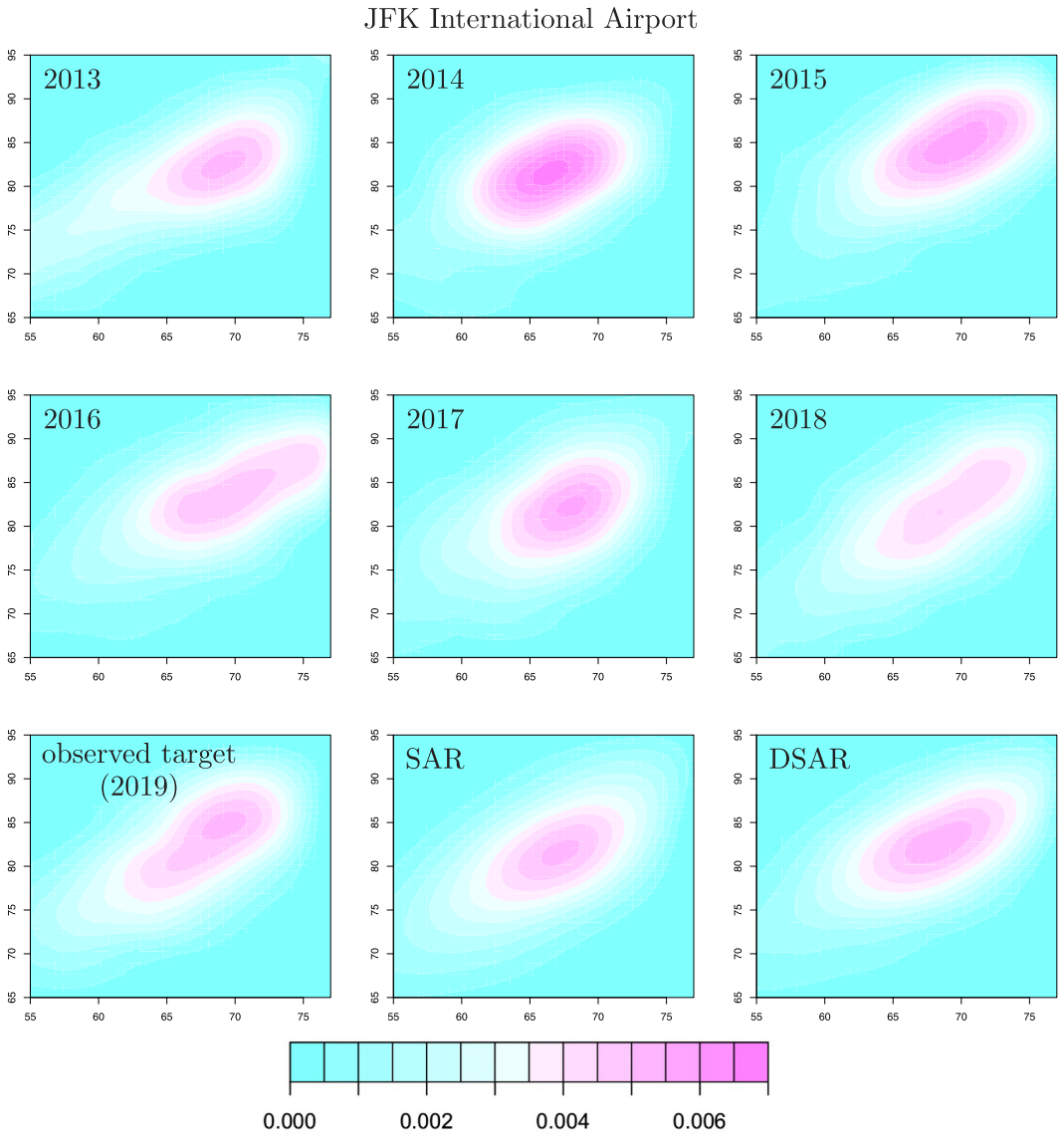

Fig. 4. Contour plots of observed and predicted two-dimensional density functions for the distributional time series of temperatures as recorded at JFK. The top six panels show the observed density functions in the training set. The bottom left panels show the observed distribution for 2019 (left); the predicted density using SAR (middle), with Fisher–Rao distance between predicted and observed of 0.147; and the predicted density using DSAR, with Fisher–Rao distance 0.186.

specifically, for $K = \lfloor \rho n \rfloor$, where $\rho \in (0, 1)$ and $k = 1, \ldots, K$, we use $\{x_{K-k+1}, \ldots, x_{n-k}\}$ to train models SAR or DSAR and to obtain a prediction $\widehat{x}_{n-k+1}$ at time $n - k + 1$. We select the order p so as to minimize $\sum_{k=1}^{K} d_{\mathcal{S}}(\widehat{x}_{n-k+1}, x_{n-k+1})/K$.

In our simulation, we compare the proposed models with vector autoregressive models (VAR) for Euclidean valued time series that ignore the spherical structure of the data. The VAR model is conveniently implemented through the R package "vars" (Pfaff, 2008). For each experiment, a spherical time series $\{x_i\}_{i=1}^{51}$ is sampled from model (10). We use $\{x_i\}_{i=1}^{50}$ as training data to predict x_{51} and repeat this 300 times, then calculate the average Fisher–Rao distance between the prediction $\hat{x}_{51}$ and the observed value x_{51}. The results are summarized in Table 1.

We find that the proposed SAR is the best method for case (i), the variant DSAR is the best method for case (ii) and both SAR and DSAR consistently outperform VAR. The simulation results thus suggest that taking the geometric structure of spherical time series into account is worthwhile and leads to better results for prediction by capturing the underlying structure.

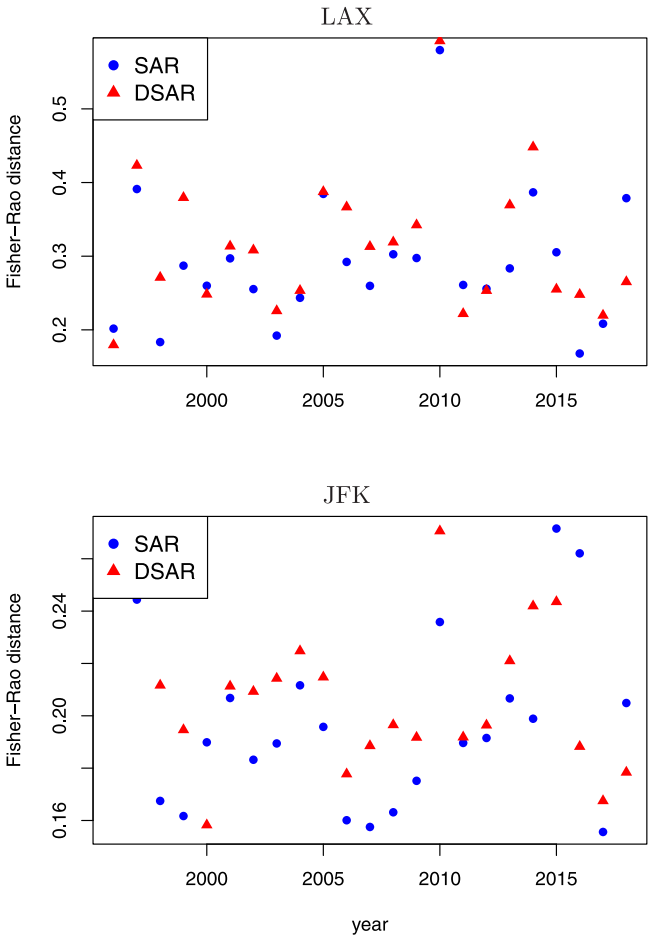

Fig. 5. Fisher–Rao distances between observed and fitted densities for each year for distributional time series of two-dimensional temperature distributions.

Table 1
Simulation results for model (10), where SAR, DSAR are defined in Eq. (10).

$(\alpha_1, \alpha_2, \alpha_3)$	Case (i)			Case (ii)		
	SAR	DSAR	VAR	SAR	DSAR	VAR
(0.5, 0, 0)	0.684	0.743	0.973	1.350	1.160	1.438
(0.6, −0.3, 0)	0.741	0.829	1.049	1.340	1.151	1.419
(0.3, −0.1, 0.4)	0.658	0.722	0.943	1.367	1.310	1.511

8. Discussion

While both compositional and distributional time series can be represented as spherical time series, such time series also arise for directional data (Mardia, 2014). Vector time series may also be represented with a spherical component if one is primarily interested in the directions of the vectors over time and less in their length, via polar coordinates. All of this adds to the motivation to study spherical time series. There is not much in terms of methodology available at this time and there is clearly room for the development of advanced time series models. In this paper we attempt to address this dearth of methodology by developing an autoregressive spherical model. We propose to represent rotation operators on spheres by skew-symmetric operators that can be viewed as elements of a Hilbert space so that linear operations become available. Other approaches may also be of future interest. Our goal is to provide a first modeling approach as a baseline with which future approaches can be compared.

It is of course possible to use alternative metrics for both compositional and distributional time series. For compositional data, a classical alternative is the Aitchison geometry (Aitchison, 1986), which also has been extended to distributional data (Hron et al., 2016). However, in applications to compositional data this approach does not work if

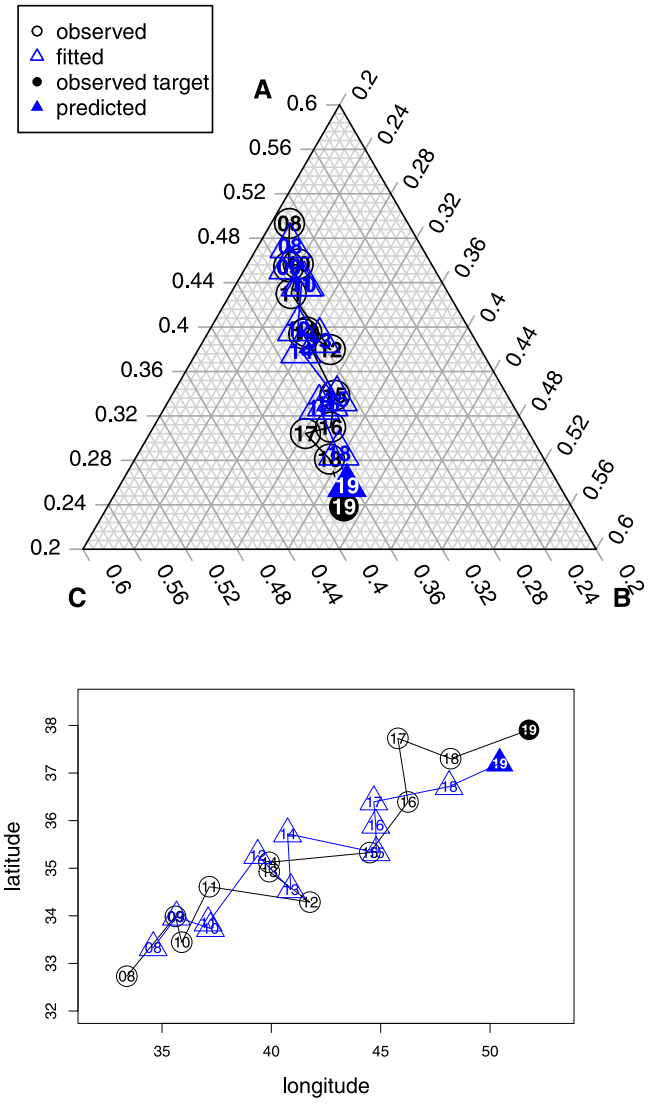

Fig. 6. Observed (circles), fitted (triangles) and prediction target (19, circle is observed and triangle is predicted) for the US energy sources compositional time series data, when fitting model DSAR. The numbers 08, . . . , 09, 10, . . . , 19 indicate the years from 2008 to 2019. Top panel: Ternary plot reflecting the compositional nature of the data; here the corner A represents coal or petroleum; B represents natural gas; and C represents nuclear and renewables. Bottom panel: The compositional time series and predictions shown in spherical coordinates. The Fisher–Rao distance between predicted and observed compositions for 2019 is 0.0223.

some of the component fractions are zero, requiring arbitrary adjustments; it also requires the arbitrary selection of a baseline component; the spherical approach does not face these difficulties (Scealy and Welsh, 2014).

For distributional time series an obvious alternative is to consider the space of distributions equipped with the Wasserstein metric (Villani, 2009) that is connected with optimal transport. When adopting this metric, the time series is not spherical and needs to be modeled in the Wasserstein manifold, where one can use tangent bundles (Chen et al., 2021; Zhang et al., 2022) or an intrinsic optimal transport approach (Zhu and Müller, 2021). However when dealing with the Wasserstein space for multivariate distributions one faces major hurdles in both theory and computation. In contrast, the Fisher–Rao metric that we consider here allows seamless extensions to any dimension. When the distributions are unknown, they need to be estimated and density estimation in higher dimensions is subject to the curse of dimensionality. This can be counteracted by assuming that the number of data from which each of the densities is estimated is large relative to the number of densities that one considers in the model.

Further in-depth comparisons of the various possible approaches to distributional and compositional time series will need to await future research. Beyond these two signature applications, autoregressive models for spherical time series provide a useful tool for directional time series and other situations where one has a natural representation of data on

a finite- or infinite-dimensional sphere. Another area of future research will be the development of other time series approaches for such data that extend autoregressive models to more complex models for time series such as GARCH models or to the frequency domain.

Finally, the spherical regression models that we have proposed here are also applicable for the case of a multiple regression in a non-time series context, for situations where both predictors and responses are spheres. In this case one has n i.i.d. pairs $(X_{i1}, \ldots, X_{im}, Y_i) \in \mathcal{S}$ and aims to model and obtain fits for the regression relation $E(Y|X_1, \ldots, X_m)$. To our knowledge, such multiple spherical regression models have not been studied yet.

9. Technical details

9.1. Proof of Theorem 1

By construction, $\{u_1, u_2\}$ is an orthonormal basis for $\text{span}\{x_1, x_2\}$. To rotate a point $x \in \mathcal{H}$, we decompose x into two parts

$$x = \text{Proj}_{u_1,u_2} x + (x - \text{Proj}_{u_1,u_2} x),$$

where $\text{Proj}_{u_1,u_2} = u_1 \otimes u_1 + u_2 \otimes u_2$ is the projection onto $\text{span}\{u_1, u_2\}$. The part that is orthogonal to $\text{span}\{u_1, u_2\}$ stays the same before and after rotating. Observe that for any $y \in \mathcal{H}$

$$\begin{aligned} Q^2(y) &= Q(Q(y)) \\ &= Q(\langle u_1, y\rangle u_2 - \langle u_2, y\rangle u_1) \\ &= -(\langle u_1, y\rangle u_1 + \langle u_2, y\rangle u_2) \\ &= -(u_1 \otimes u_1 + u_2 \otimes u_2)y \end{aligned}$$

which entails that $Q^2 = -\text{Proj}_{u_1,u_2}$. For the part within $\text{span}\{u_1, u_2\}$, we can apply the two dimensional rotations and map the result back to $\mathcal{H}$, i.e., $\text{Proj}_{u_1,u_2} x$ is rotated to

$$\begin{aligned} &(\cos(\vartheta)\langle u_1, x\rangle - \sin(\vartheta)\langle u_2, x\rangle)u_1 + (\sin(\vartheta)\langle u_1, x\rangle + \cos(\vartheta)\langle u_2, x\rangle)u_2 \\ &= \cos(\vartheta)(u_1 \otimes u_1 + u_2 \otimes u_2)x + \sin(\vartheta)(u_1 \otimes u_2 - u_2 \otimes u_1)x \\ &= (\cos(\vartheta)(-Q^2) + \sin(\vartheta)Q)x, \end{aligned}$$

Using $Q^2 = -\text{Proj}_{u_1,u_2}$ and rearranging the terms, the point x is seen to be rotated to

$$(\cos(\vartheta)(-Q^2) + \sin(\vartheta)Q)x + (I + Q^2)x = \left(I + \sin(\vartheta)Q + (1 - \cos(\vartheta))Q^2\right)x.$$

To show identity (2), observe that for any $y \in \mathcal{H}$,

$$\begin{aligned} Q^3(y) &= Q(Q^2(y)) \\ &= -Q(\langle u_1, y\rangle u_1 + \langle u_2, y\rangle u_2) \\ &= -(\langle u_1, y\rangle u_2 - \langle u_2, y\rangle u_1) \\ &= -Qy, \end{aligned}$$

which entails that $Q^3 = -Q$. By adopting the idea for rotations in $\text{SO}(n)$ (Gallier and Xu, 2003), we expand the exponential map and apply $Q^3 = -Q$ repeatedly,

$$\begin{aligned} \exp(\vartheta Q) &= \sum_{l=0}^{\infty} \frac{\vartheta^l Q^l}{l!} \\ &= I + \left(\sum_{l=1}^{\infty} \frac{\vartheta^{2l-1}(-1)^{l-1}}{(2l-1)!}\right) Q + \left(\sum_{l=1}^{\infty} \frac{\vartheta^{2l}(-1)^{l-1}}{(2l)!}\right) Q^2 \\ &= I + \sin(\vartheta)Q + (1 - \cos(\vartheta))Q^2. \end{aligned}$$

We then prove that $\exp(\vartheta Q)$ is orthogonal. By applying the identities $\langle Qy_1, y_2\rangle = -\langle y_1, Qy_2\rangle$ and $Q^3 = -Q$, some straight-forward algebra show that

$$\begin{aligned} &\langle \exp(\vartheta Q)y_1, \exp(\vartheta Q)y_2\rangle \\ &= \langle y_1, y_2\rangle + (1 - \cos(\vartheta))(\langle y_1, Q^2 y_2\rangle + \langle Q^2 y_1, y_2\rangle) \\ &\quad + \sin^2(\vartheta)\langle Qy_1, Qy_2\rangle + (1 - \cos(\vartheta))^2\langle Q^2 y_1, Q^2 y_2\rangle \\ &= \langle y_1, y_2\rangle - 2(1 - \cos(\vartheta))\langle Qy_1, Qy_2\rangle \\ &\quad + (\sin^2(\vartheta) + (1 - \cos(\vartheta))^2)\langle Qy_1, Qy_2\rangle \\ &= \langle y_1, y_2\rangle. \end{aligned}$$

In addition, for any $x \in \text{span}\{u_1, u_2\} \cap \mathcal{S}$, we can write $x = c_1u_1 + c_2u_2$ such that $c_1^2 + c_2^2 = 1$. Since $Qx = c_1u_2 - c_2u_1$ and $Q^2x = -c_2u_2 - c_1u_1$, we have

$$\begin{aligned}&\langle \exp(\vartheta Q)x, x\rangle\\ =&\langle x + \sin(\vartheta)(c_1u_2 - c_2u_1) - (1 - \cos(\vartheta))x, x\rangle\\ =&\cos(\vartheta).\end{aligned}$$

This shows that $\arccos(\langle \exp(\vartheta Q)x, x\rangle) = \vartheta$. Finally, since $Qy = 0$ for any y perpendicular to $\text{span}\{u_1, u_2\}$, we have $\exp(\vartheta Q)y = y$.

9.2. Proof of Theorem 3

Hereafter, we drop the subscript $\mathcal{H} \otimes \mathcal{H}$ in $\langle \cdot, \cdot \rangle_{\mathcal{H}\otimes\mathcal{H}}$ for simplicity. Denote $\widetilde{\boldsymbol{\lambda}} = (\widetilde{\lambda}_0, \widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_p)$, where for $k = 0, 1, 2, \ldots, p$, $\widetilde{\lambda}_k$ is defined as

$$\widetilde{\lambda}_k = \frac{1}{n}\sum_{t=1}^{n}\langle R_t - \mu_R, R_{t+k} - \mu_R\rangle.$$

To prove the theorem, we show the asymptotic normality of $\widetilde{\boldsymbol{\lambda}}$ and then we show that $|\widehat{\lambda}_k - \widetilde{\lambda}_k| = o_p(1)$ for any $k = 0, 1, 2, \ldots, p$. For convenience, we define $\psi_i = 0$ if $i < 0$ and write

$$R_t - \mu_R = \sum_{i=-\infty}^{\infty}\psi_i\varepsilon_{t-i}$$

Next, we compute the asymptotic covariance of $\widetilde{\boldsymbol{\lambda}}$. To this end, we observe

$$E[\langle \varepsilon_i, \varepsilon_j\rangle\langle \varepsilon_k, \varepsilon_l\rangle] = \begin{cases} E[\langle \varepsilon_1, \varepsilon_1\rangle^2], & i = j = k = l,\\ (E[\langle \varepsilon_1, \varepsilon_1\rangle])^2, & i = j \neq k = l,\\ E[\langle \varepsilon_1, \varepsilon_2\rangle^2], & i = k \neq j = l,\\ E[\langle \varepsilon_1, \varepsilon_2\rangle^2], & i = l \neq j = k,\\ 0, & \text{otherwise.}\end{cases}$$

Some algebra leads to

$$E[\langle R_t - \mu_R, R_{t+u} - \mu_R\rangle\langle R_{t+h} - \mu_R, R_{t+h+v} - \mu_R\rangle] = \Gamma_{u,v}^h,$$

where

$$\begin{aligned}\Gamma_{u,v}^h =& \left(E[\langle \varepsilon_1, \varepsilon_1\rangle^2] - (E[\langle \varepsilon_1, \varepsilon_1\rangle])^2 - 2E[\langle \varepsilon_1, \varepsilon_2\rangle^2]\right)\sum_i \psi_i\psi_{i+u}\psi_{i+h}\psi_{i+h+v}\\ &+(E[\langle \varepsilon_1, \varepsilon_1\rangle])^2\kappa(u)\kappa(v) + E[\langle \varepsilon_1, \varepsilon_2\rangle^2]\left(\kappa(h)\kappa(h+v-u) + \kappa(h+v)\kappa(h-u)\right).\end{aligned}$$

The absolute summability of $\{\Gamma_{u,v}^h\}_{h\in\mathbb{N}}$ is guaranteed by the absolute summability of $\{\psi_i\}$. For $h = s - t$,

$$\begin{aligned}\text{cov}(\widetilde{\lambda}_u, \widetilde{\lambda}_v) =& \frac{1}{n^2}\sum_{s,t=1}^{n}E[\langle R_t - \mu_R, R_{t+u} - \mu_R\rangle\langle R_s - \mu_R, R_{s+v} - \mu_R\rangle]\\ =& \frac{1}{n^2}\sum_{|h|<n}(n - |h|)\Gamma_{u,v}^h.\end{aligned}$$

Thus, let $V = \lim_{n\to n} n\times\text{cov}(\widetilde{\boldsymbol{\lambda}})$, it can be seen that $V_{u,v} := \lim_{n\to\infty} n\times\text{cov}(\widetilde{\lambda}_u, \widetilde{\lambda}_v) = \sum_{h=-\infty}^{\infty}\Gamma_{u,v}^h$. To derive the asymptotic distribution of $\widetilde{\boldsymbol{\lambda}}$, we follow the idea in Chapter 7 of Brockwell and Davis (1991) by defining $T_t^m = \sum_{i=-m}^{m}\psi_i\varepsilon_{t-i}$, $\widetilde{\lambda}_k^m = \sum_{t=1}^{n}\langle T_t^m, T_{t+k}^m\rangle/n$ and $\widetilde{\boldsymbol{\lambda}}^m = (\widetilde{\lambda}_0^m, \widetilde{\lambda}_1^m, \ldots, \widetilde{\lambda}_p^m)^T$. Notice that for any k, $\{\langle T_t^m, T_{t+k}^m\rangle\}_{t=1}^n$ is a $(2m + p)$-dependent sequence. Applying the CLT for m-dependent data leads to

$$\sqrt{n}\left(\widetilde{\boldsymbol{\lambda}}^m - \boldsymbol{\lambda}^m\right) \to^d N(0, V^m),$$

where $\boldsymbol{\lambda}^m = E[\widetilde{\boldsymbol{\lambda}}^m]$ and $V^m := \lim_{n\to\infty} n \times \text{cov}(\widetilde{\boldsymbol{\lambda}}^m)$. It is clear that $V^m \to V$ as $m \to \infty$. Next, we show that

$$\lim_{m\to\infty}\limsup_{n\to\infty} P\left(\left|\sqrt{n}(\widetilde{\boldsymbol{\lambda}}^m - \boldsymbol{\lambda}^m - \widetilde{\boldsymbol{\lambda}} + \boldsymbol{\lambda})\right| > \epsilon\right) \to 0, \tag{11}$$

where $|\cdot|$ is the L^2-distance. Notice that the above probability is bounded by $\sum_{k=0}^{p} P(\sqrt{n}|\widetilde{\lambda}_k^m - \lambda_k^m - \widetilde{\lambda}_k + \lambda_k| > \epsilon/\sqrt{p+1})$ and

$$P(\sqrt{n}|\widetilde{\lambda}_k^m - \lambda_k^m - \widetilde{\lambda}_k + \lambda_k| > \epsilon/\sqrt{p+1}) \le \frac{p+1}{\epsilon^2} n \left(\mathrm{var}(\widetilde{\lambda}_k^m) + \mathrm{var}(\widetilde{\lambda}_k) - 2\mathrm{cov}(\widetilde{\lambda}_k^m, \widetilde{\lambda}_k)\right).$$

Since $\lim_{m\to\infty} \lim_{n\to\infty} n(\mathrm{var}(\widetilde{\lambda}_k^m) + \mathrm{var}(\widetilde{\lambda}_k) - 2\mathrm{cov}(\widetilde{\lambda}_k^m, \widetilde{\lambda}_k)) = (V_{k,k} + V_{k,k} - 2V_{k,k}) = 0$, Eq. (11) holds, then it follows from Proposition 6.3.9 of Brockwell and Davis (1991) that

$$\sqrt{n}\left(\widetilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}\right) \to^d N(0, V).$$

Finally, we show that $\widetilde{\boldsymbol{\lambda}} - \widehat{\boldsymbol{\lambda}} = o_p(1)$. It suffices to show $\sqrt{n}(\widetilde{\lambda}_k - \widehat{\lambda}_k) = o_p(1)$ for any $k = 0, 1, \ldots, p$. Some algebra shows that

$$\widetilde{\lambda}_k - \widehat{\lambda}_k = O_p\left(\langle \widehat{\mu}_R - \mu_R, \widehat{\mu}_R - \mu_R\rangle\right).$$

Since $\{\kappa(u)\}_{u\in\mathbb{N}}$ is absolutely summable,

$$\begin{aligned} E(\langle \widehat{\mu}_R - \widehat{\mu}_R, \widehat{\mu}_R - \mu_R\rangle) &= E\left(\frac{1}{n^2}\sum_{s,t=1}^{n}\langle R_s - \mu_R, R_t - \mu_R\rangle\right) \\ &= \frac{1}{n} O\left(\kappa(0) + \sum_{k=1}^{n-1} 2(1-k/n)\kappa(k)\right) \\ &= O\left(\frac{1}{n}\right). \end{aligned}$$

Also, as $\{\Gamma_{u,v}^h\}_{u\in\mathbb{N}}$ is absolutely summable, we have

$$\begin{aligned} &\mathrm{var}(\langle \widehat{\mu}_R - \widehat{\mu}_R, \widehat{\mu}_R - \mu_R\rangle) \\ &= \frac{1}{n^4}\sum_{s_1,t_1,s_2,t_2} \mathrm{cov}\left(\langle R_{s_1} - \mu_R, R_{t_1} - \mu_R\rangle, \langle R_{s_2} - \mu_R, R_{t_2} - \mu_R\rangle\right) \\ &= \frac{1}{n} O\left(\frac{1}{n^3}\sum_{s_1,s_2,t_2}\left(\sum_{t_1} \Gamma_{|t_1-s_1|,|t_2-s_2|}^{|s_2-s1|}\right)\right) \\ &= O(1/n). \end{aligned}$$

We conclude that $\langle \widehat{\mu}_R - \mu_R, \widehat{\mu}_R - \mu_R\rangle = o_p(1)$ and $\sqrt{n}(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \to^d N(0, V)$.

References

Aitchison, J., 1986. The Statistical Analysis of Compositional Data. Chapman & Hall, Ltd.
Bauer, M., Bruveris, M., Michor, P.W., 2016. Uniqueness of the Fisher–Rao metric on the space of smooth densities. Bull. Lond. Math. Soc. 48 (3), 499–506.
Bhatia, A., Katz, J., 2021. Why we are experiencing so many unusually hot summer nights. N.Y. Times September 16, A12.
Bosq, D., 2000. Linear Processes in Function Spaces: Theory and Applications. Springer-Verlag, New York.
Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods. Springer-Verlag, New York.
Burago, D., Burago, Y., Ivanov, S., 2001. A Course in Metric Geometry. American Mathematical Society, Providence, RI.
Chang, T., 1986. Spherical regression. Ann. Statist. 14 (3), 907–924.
Chang, T., 1989. Spherical regression with errors in variables. Ann. Statist. 17 (1), 293–306.
Chen, Y., Lin, Z., Müller, H.-G., 2021. Wasserstein regression. J. Amer. Statist. Assoc. 1–14.
Dai, X., 2022. Statistical inference on the Hilbert sphere with application to random densities. Electron. J. Stat. 16 (1), 700–736.
Dai, X., Müller, H.G., 2018. Principal component analysis for functional data on Riemannian manifolds and spheres. Ann. Stat. 46, 3334–3361.
Downs, T.D., 2003. Spherical regression. Biometrika 90 (3), 655–668.
Fan, J., Yao, Q., 2003. Nonlinear Time Series: Nonparametric and Parametric Methods. Springer-Verlag New York.
Fan, J., Yao, Q., 2017. The Elements of Financial Econometrics. Cambridge University Press.
Felicísimo, Á.M., Muñoz, J., González-Solis, J., 2008. Ocean surface winds drive dynamics of transoceanic aerial movements. PLoS One 3 (8), e2928.
Friedrich, T., 1991. Die Fisher-Information und symplektische Strukturen. Math. Nachr. 153 (1), 273–296.
Gajardo, A., Carroll, C., Chen, Y., Dai, X., Fan, J., Hadjipantelis, P.Z., Han, K., Ji, H., Müller, H.-G., Wang, J.-L., 2021. fdapace: Functional data analysis and empirical dynamics. R package version 0.5.7, https://CRAN.R-project.org/package=fdapace.
Gallier, J., Xu, D., 2003. Computing exponentials of skew-symmetric matrices and logarithms of orthogonal matrices. Int. J. Robot. Autom. 18 (1), 10–20.
Hron, K., Menafoglio, A., Templ, M., Hruzova, K., Filzmoser, P., 2016. Simplicial principal component analysis for density functions in Bayes spaces. Comput. Stat. Data Anal. 94, 330–350.
Kim, P.T., 1998. Deconvolution density estimation on SO(N). Ann. Statist. 26 (3), 1083–1102.
Mardia, K.V., 2014. Statistics of Directional Data. Academic Press.
Martin, M.H., 1932. On infinite orthogonal matrices. Amer. J. Math. 54 (3), 579–631.
Marzio, M.D., Panzera, A., Taylor, C.C., 2019. Nonparametric rotations for sphere-sphere regression. J. Amer. Statist. Assoc. 114 (525), 466–476.
Müller, H.G., 2016. Peter hall, functional data analysis and random objects. Ann. Stat. 44, 1867–1887.

Pegoraro, M., Beraha, M., 2022. Projected statistical methods for distributional data on the real line with the Wasserstein metric. J. Mach. Learn. Res. 23, 1–59.
Petersen, A., Liu, X., Divani, A.A., 2021. Wasserstein F-tests and confidence bands for the Fréchet regression of density response curves. Ann. Stat. 49 (1), 590–611.
Petersen, A., Müller, H.G., 2016. Functional data analysis for density functions by transformation to a Hilbert space. Ann. Stat. 44 (1), 183–218.
Pfaff, B., 2008. VAR, SVAR and SVEC models: Implementation within R package vars. J. Stat. Softw. 27 (4).
Rosenthal, M., Wu, W., Klassen, E., Srivastava, A., 2014. Spherical regression models using projective linear transformations. J. Amer. Statist. Assoc. 109 (508), 1615–1624.
Scealy, J., Welsh, A., 2011. Regression for compositional data by using distributions defined on the hypersphere. J. R. Stat. Soc.: Ser. B (Stat. Methodol.) 73 (3), 351–375.
Scealy, J., Welsh, A., 2014. Colours and cocktails: Compositional data analysis. Aust. N. Z. J. Stat. 56 (2), 145–169.
Shi, X., Li, X., Cai, T., 2021. Spherical regression under mismatch corruption with application to automated knowledge translation. J. Amer. Statist. Assoc. 116 (536), 1953–1964.
Villani, C., 2009. Optimal Transport: Old and New. Springer.
Zhang, C., Kokoszka, P., Petersen, A., 2022. Wasserstein autoregressive models for density time series. J. Time Series Anal. 43 (1), 30–52.
Zhu, C., Müller, H.G., 2021. Autoregressive optimal transport models. arXiv preprint arXiv:2105.05439.