The Annals of Statistics
2024, Vol. 52, No. 2, 757–792
https://doi.org/10.1214/24-AOS2368

METRIC STATISTICS: EXPLORATION AND INFERENCE FOR RANDOM OBJECTS WITH DISTANCE PROFILES

BY PAROMITA DUBEY[1,a], YAQING CHEN[2,b] AND HANS-GEORG MÜLLER[3,c]

[1]*Department of Data Sciences and Operations, Marshall School of Business, University of Southern California,* [a]*paromita@marshall.usc.edu*

[2]*Department of Statistics, Rutgers University,* [b]*yqchen@stat.rutgers.edu*

[3]*Department of Statistics, University of California, Davis,* [c]*hgmueller@ucdavis.edu*

This article provides an overview on the statistical modeling of complex data as increasingly encountered in modern data analysis. It is argued that such data can often be described as elements of a metric space that satisfies certain structural conditions and features a probability measure. We refer to the random elements of such spaces as random objects and to the emerging field that deals with their statistical analysis as metric statistics. Metric statistics provides methodology, theory and visualization tools for the statistical description, quantification of variation, centrality and quantiles, regression and inference for populations of random objects, inferring these quantities from available data and samples. In addition to a brief review of current concepts, we focus on distance profiles as a major tool for object data in conjunction with the pairwise Wasserstein transports of the underlying one-dimensional distance distributions. These pairwise transports lead to the definition of intuitive and interpretable notions of transport ranks and transport quantiles as well as two-sample inference. An associated profile metric complements the original metric of the object space and may reveal important features of the object data in data analysis. We demonstrate these tools for the analysis of complex data through various examples and visualizations.

1. Introduction. We delineate in this article an emerging field of statistics that provides models, methods and theory for complex data situated in metric spaces (Ω, d) with a metric d. We refer to this field as *metric statistics*. Throughout it is assumed that the metric spaces where the data are situated are separable and endowed with a probability measure P. We refer to random variables taking values in such metric spaces as *random objects*, adopting the name from a previous review and perspective (Müller (2016)).

The motivation to address the challenges posed by non-Euclidean data and to study common features of such data and techniques that are applicable across many metric spaces comes from data analysis, where increasingly complex data objects are encountered. Statistical analysis means that the emphasis is on statistical methods that evolved from and have counterparts in classical Euclidean statistics, are interpretable rather than black-box approaches, and are amenable to uncertainty quantification and inference. The need for such methodology has not gone unnoticed and over the last two decades various groups of statisticians have come up with interesting and important ideas about the handling of such data. This includes object-oriented data analysis with roots in statistics for manifold-valued data, shape analysis and geometric statistics and related ideas for visualization and modeling (Dryden, Koloydenko and Zhou (2009), Huckemann and Eltzner (2021), Marron and Dryden (2021), Wang and Marron (2007)), and also symbolic data analysis, where various subproblems have been emphasized such as data that consist of intervals (Billard and Diday (2003)).

Received November 2023; revised February 2024.

Key words and phrases. Distributional data, Fréchet mean, Fréchet regression, functional data analysis, metric variance, profile metric, transport rank, transport quantile, visualization, Wasserstein metric.

A distinctive feature of metric statistics that differentiates it from classical as well as geometric statistics is the nonreliance on local or global Euclidean or manifold structure. While for some spaces local linearizations may exist, as exemplified by one-dimensional distributional data with the 2-Wasserstein metric, where one can use the pseudo Riemannian structure to define L^2 tangent spaces (Bigot et al. (2017), Chen, Lin and Müller (2023)), these are often only of limited utility; for example, inverse maps from the linear spaces back to the metric space usually are not well defined on the entire linear approximation space. The same holds for linear embeddings into a subset of a Hilbert space obtained through kernel maps (Schoenberg (1938), Sejdinovic et al. (2013)), although there exist specific invertible maps to a Hilbert space for special nonlinear spaces, which however induce metric distortions (Petersen and Müller (2016a)). The lack of Euclidean structure in general metric spaces poses challenging problems for statistical theory, methodology and data analysis of random objects and essentially requires to rethink basic notions of mean, variation, regression, inference and other key statistical techniques. The overall goal is to arrive at a principled, theory-supported and comprehensive toolkit for the analysis of samples of random objects.

After a brief review, we focus here on distance profiles, a basic tool that assigns a one-dimensional distribution to each element of the underlying metric space (Ω, d). Distance profiles are the distributions of the distances of each element to a random object in the space Ω and are determined by the underlying probability measure P on Ω. As we will show, distance profiles not only reflect but indeed characterize P under some regularity assumptions. In all of the following, we will assume that one has a sample of i.i.d. random objects drawn from P. Empirical estimates for the distance profiles are then simply obtained using the empirical distribution of the distances of any given element of Ω to all other elements, either to all elements in the population or in the empirical version to the other sample elements. We will illustrate this idea also for the simple and familiar special case of Euclidean data; in all scenarios, distance profiles always correspond to one-dimensional distributions.

Distance profiles have multiple applications that we explore in this article. First of all, they aid the geometric exploration of random objects in (Ω, d) under the measure P. Second, since distance profiles are always one-dimensional distributions, we can define a new dissimilarity measure on Ω by adopting a metric on the space of one-dimensional distributions, which is then applied to the distributional distance of the distance profiles of the two elements. This dissimilarity measure depends on both the original metric in the space (Ω, d) as well on the metric adopted for one-dimensional distributions, where here we adopt the 2-Wasserstein metric as the metric in the distributional space of distance profiles.

Third, the pairwise transports that result from adopting the 2-Wasserstein metric for the space of distance profiles make it possible to define novel notions of transport centrality and associated transport ranks. These serve to quantify the centrality of objects and provide the basis for a partial ordering of random objects and resulting visualizations. Fourth, transport centrality and transport ranks can be harnessed to define transport quantiles as the set of elements of Ω with transport ranks such that the elements with higher ranks have a probability mass no less than the prespecified quantile level.

Fifth, we demonstrate how distance profiles across two samples can be used to test whether the probability measures that generate the samples are identical. This relies on the fact that distance profiles characterize the underlying probability measures; we note that related ideas on inference based on distance profiles as those presented in Section 6 below, resulting from seemingly independent work, were recently published (Wang et al. (2023)).

Distance profiles thus emerge as a powerful tool to characterize random objects. As we show in the following Section 2, they are natural extensions of some basic ideas of how to quantify the variation of random objects. Section 2 contains a brief review of some of the basic concepts of metric statistics, including Fréchet and transport regression. Distance profiles and how they give rise to transport ranks and quantiles, as well as notions such as most

central points and the properties of these concepts, will be the theme of Section 3 and Section 4. Further connections to applications for inference, specifically distance profile based inference, will be discussed in Section 6, followed by simulation studies and applications to age-at-death distributions of human mortality, U.S. energy generation data and functional connectivity networks based on fMRI data in Section 7. We conclude with a discussion on the choice of metrics and other topics in Section 8. Auxiliary results and proofs as well as additional simulations and data examples are provided in the Supplementary Material (Dubey, Chen and Müller (2024)).

2. Review of basic notions for samples of random objects. Random objects encompass the usual random variables that take values in spaces $\mathbb{R}^p$ as encountered in classical statistics and also random functions in Hilbert spaces L^2, which is the realm of functional data analysis and where one still has linear structures, inner products and linear operators (Hsing and Eubank (2015), Wang, Chiou and Müller (2016)). Other well-studied classes of random objects are data on Riemannian manifolds, notably spheres, which also appear in shape analysis (Dryden and Mardia (2016), Jung, Dryden and Marron (2012)) and where surprising smeariness results were obtained in recent developments on the limit theory for Fréchet means (Eltzner and Huckemann (2019)).

A recently emerging subarea of metric statistics is distributional data analysis, where the atoms of a sample are distributions. These may be directly observed or more commonly indirectly through the data that each distribution generates. In earlier approaches, samples of distributions were treated as functional data (Kneip and Utikal (2001)), but while density or distribution functions can be considered as elements of the function space L^2, this approach is suboptimal since distributional objects lie on a constrained submanifold; for example, densities are nonnegative and integrate to 1. Taking these constraints fully into account for statistical analysis motivates distributional data analysis (Ghosal et al. (2023), Matabuena et al. (2021), Petersen, Zhang and Kokoszka (2022)). Linearization approaches for distributional data include the Bayes space transformation (Hron et al. (2016)), which is based on the Aitchison geometry (Aitchison (1986)), however, does not yield a 1:1 map, and a class of 1:1 transformations to linear spaces that includes the log quantile density (lqd) and log hazard transformations (Petersen and Müller (2016a)). More recent approaches have used local linearizations through the geometry of the Wasserstein manifold (Chen, Lin and Müller (2023), Pegoraro and Beraha (2022)) and fully intrinsic optimal transport models that do not rely on any ambient L^2 space (Ghodrati and Panaretos (2023), Zhu and Müller (2023a)). In distributional data analysis, the metric space Ω is the space of distributions, which are often assumed to have a finite domain and to be continuous, and d is an appropriate metric. For statistical analysis in the case of one-dimensional distributions the 2-Wasserstein metric has become popular, not least due to its practical appeal in data analysis (Bolstad et al. (2003)). For probability measures μ, ν with distribution functions F_μ, F_ν, the 2-Wasserstein distance is given simply as the L^2 distance of the quantile functions,

$$d_W(\mu,\nu) = \left(\int_0^1 \left[F_\mu^{-1}(u) - F_\nu^{-1}(u)\right]^2 \mathrm{d}u\right)^{1/2} = d_{L^2}(F_\mu^{-1}, F_\nu^{-1}). \tag{1}$$

For the case of multivariate distributions, quantile functions do not exist and the 2-Wasserstein metric is more directly tied to optimal transport in the Monge–Kantorovich transportation problem, where the Kantorovich version (Kantorovitch (1958)) is

$$d_W^2(\mu,\nu) = \inf_{\wp \in \mathscr{P}(\mu,\nu)} \mathbb{E}_{(X,Y)\sim\wp}\|X-Y\|^2. \tag{2}$$

Here, X, Y are random variables in $\mathbb{R}^d$, μ, ν are probability measures supported on a set $D \subset \mathbb{R}^d$, and $\mathscr{P}(\mu,\nu)$ is the space of joint probability measures on $D \times D$ with marginals μ

and ν. If the probability distributions are absolutely continuous, this is equivalent to finding the optimal transport in Monge's version

$$T^*(\mu,\nu) = \operatorname*{arginf}_{T:T_\#\mu=\nu} \mathbb{E}_{X\sim\mu}\|X - T(X)\|^2. \tag{3}$$

Here, the infimum is taken over all push-forward Borel maps $D \to D$ that map μ to ν. The push-forward map T applied to a probability measure ν_1 on D yields $\nu_2 = T_\#\nu_1$, defined as the measure with $\nu_2(A) = \nu_1(T^{-1}(A))$ for any measurable set $A \subset D$. If it exists, this minimizer is the optimal transport map and the minimizing value is the Wasserstein metric d_W, which coincides with the definition in (1) for the special case of univariate distributions.

In the multivariate case, this minimization problem is computationally challenging and often replaced by a relaxed version, for example the Sinkhorn minimization (Cuturi (2013)), but then depends critically on regularization parameters. The statistical motivation to use the Wasserstein metric for multivariate distributional data is also less compelling than for the one-dimensional case. Therefore, it makes sense to use a simpler metric for this case. Options include the sliced Wasserstein metric (Kolouri et al. (2019)) or the Fisher–Rao metric, which does not have quite the appeal of the Wasserstein metric with its connection to optimal mass transport, but is easy to compute in any dimension when densities exist, as it is the geodesic distance for the square roots of densities. These square roots are situated on the Hilbert sphere, whence for measures ν_1, ν_2 with densities f_{ν_1}, f_{ν_2} the metric is

$$d_{\mathrm{FR}}(\nu_1,\nu_2) = \arccos\left\{\int \sqrt{f_{\nu_1}(x) f_{\nu_2}(x)}\,\mathrm{d}x\right\}.$$

When adopting the Fisher–Rao metric, multivariate distributional data and spherical data as are commonly encountered in directional data analysis can be viewed in a unified framework of spherical data, for example, in time-series analysis (Zhu and Müller (2024)). The spherical framework also encompasses compositional data when using the square root transformation for proportions (Scealy and Welsh (2014)).

Other important classes of random objects include covariance matrices and surfaces (Pigoli et al. (2014), Zemel and Panaretos (2019)), networks (Severn, Dryden and Preston (2022), Zhou and Müller (2022)) and trees (Barden, Le and Owen (2018), Garba et al. (2021), Lueg et al. (2022)), where for the BHV metric (Billera, Holmes and Vogtmann (2001)) the requisite entropy conditions for the asymptotic analysis of M-estimators were recently established (Lin and Müller (2021)). Analogous to functional analysis and linear operator theory being the basis of functional data analysis, so is metric geometry (Burago, Burago and Ivanov (2001)) the basis for metric statistics; an abbreviated introduction for statisticians can be found in Section 2 and Appendix B in Lin and Müller (2021).

We aim to find commonalities across metric spaces, unifying theory and methodology, regardless of the specific geometry of the metric space. Entropy conditions that quantify the size of the space have emerged as a key tool. The utility of entropy conditions and empirical process theory for random objects was recognized in recent work on Fréchet means/barycenters and Fréchet regression (Ahidar-Coutrix, Le Gouic and Paris (2020), Petersen and Müller (2019), Schötz (2019), Schötz (2022)). A basic and classical notion is the measure of location provided by the Fréchet mean (Fréchet (1948)) or barycenter, which is defined as minimizer of the Fréchet function $\Omega \to \mathbb{R}$,

$$V(\omega) = \mathbb{E}d^2(\omega, X). \tag{4}$$

The population/sample minimizers

$$\mu_\oplus = \operatorname*{argmin}_{\omega\in\Omega} \mathbb{E}d^2(\omega, X), \qquad \widehat{\mu}_\oplus = \operatorname*{argmin}_{\omega\in\Omega} \sum_{i=1}^n d^2(\omega, X_i) \tag{5}$$

generally form a set.

Uniqueness of Fréchet means is guaranteed in Hadamard spaces (Sturm (2003)), and for positively curved spaces depends on both the geometry of the space Ω and the probability measure P. Analogously to Fréchet means one can also consider Fréchet integrals for Ω-valued functions $X(t)$ (Petersen and Müller (2016b)). A functional scenario with Ω-valued stochastic processes widens the scope of functional data analysis (Dubey and Müller (2020a)), where the previous standard has been that the underlying processes are Euclidean, either scalar-, vector- or L^2-valued (Chen, Delicado and Müller (2017)). The special case of distribution-valued stochastic processes is of particular interest and permits a more in-depth investigation (Zhou and Müller (2023)). For general types of Ω-valued functions, the Fréchet integral provides a direct extension of the Riemann integral for $\mathbb{R}$-valued functions and in analogy to the Fréchet mean is defined as

$$\int_{\oplus} X(t)\,dt = \operatorname*{argmin}_{\omega\in\Omega} \int d^2\big(\omega, X(t)\big)\,dt.$$

This integral has proved useful in various investigations of object-valued processes (Dubey and Müller (2020a), Lin and Müller (2021), Petersen and Müller (2016b)).

Another important extension of Fréchet means is the notion of a conditional Fréchet mean $\mathbb{E}_{\oplus}(X|Z)$, where $X \in \Omega$ and $Z \in \mathbb{R}^p$, or more generally $Z \in \Omega'$ for another metric space (Ω', d'). The statistical motivation is to model complex regression relationships that involve random objects. A narrower specification is needed to make this notion useful for statistical modeling and data analysis, targeting

$$m_{\oplus}(z) = \operatorname*{argmin}_{\omega\in\Omega} \mathbb{E}\big(d^2(X,\omega)|Z=z\big).$$

Nadaraya–Watson kernel estimators for the case of manifold-to-manifold regression $\Omega' \to \Omega$ that have been previously considered (Steinke and Hein (2009), Steinke, Hein and Schölkopf (2010)) are subject to a severe version of the curse of dimensionality, unless the predictor manifold is low-dimensional, and they are also subject to substantial boundary effects. Special cases include the smoothing of covariance matrices or data on a Riemannian manifolds indexed by time using local linear estimators (Cheng and Wu (2013), Cornea et al. (2017), Yuan et al. (2012)), including versions for functional and longitudinal data analysis, where functional principal components are a primary target after the smoothing step (Dai, Lin and Müller (2021)).

In addition to local linear and other desirable smoothers for the case of low-dimensional predictors, it is also of interest to include global models that extend the classical linear multiple regression model when responses are random objects and predictors are Euclidean vectors. A general approach is Fréchet regression (Petersen and Müller (2019)) for the subproblem where $\Omega' = \mathbb{R}^p$. Observing that smoothing or global linear regression methods are weighted averages with known or computable weights, one can take the weights that correspond to the respective regression method and form a weighted Fréchet mean. The key problem is that when estimating at certain predictor levels some weights will be negative. Making use of entropy conditions for the metric space Ω leads to asymptotic convergence across all spaces that satisfy these conditions, for both local and global regression models.

Denoting the weights for a classical regression method with Euclidean predictors and scalar responses that are assigned to a predictor at level $Z \in \mathbb{R}^p$ when targeting the estimate at a fixed predictor level $z \in \mathbb{R}^p$ by $w(Z,z)$, the Fréchet regression estimator is

$$\widehat{m}_{\oplus}(z) = \operatorname*{argmin}_{\omega\in\Omega} \mathbb{E}\big(d^2(X,\omega)w(Z,z)\big).$$

There are many open problems associated with this class of estimators and object regression is a subarea in rapid development. Recent work includes a novel perspective with extensions

to other smoothing methods (Schötz (2022)), dimension reduction (Dong and Wu (2022), Virta, Lee and Li (2022), Zhang, Li and Xue (2024), Zhang, Xue and Li (2021)) and consistent predictor selection (Tucker, Wu and Müller (2023)). A recent derivation of uniform convergence over the domain of the Euclidean predictor for local linear estimators made it possible to obtain consistent time warping identification for object-valued functional data through pairwise warping comparisons and also to obtain consistent estimates for the location of extrema of functionals such as a specified eigenvalue for symmetric positive definite matrices as random objects (Chen and Müller (2022)). This result also facilitated the development of a single-index version for Fréchet regression, which enhances the flexibility of the global version of the model and includes inference for the the predictors (Bhattacharjee and Müller (2023), Ghosal, Meiring and Petersen (2023)). However, much further work is needed on inference for object regression. Another class of object regression models that is of potential interest but not sufficiently explored is transport regression (Zhu and Müller (2023b)), which was initially developed for distributional data, where one can introduce a transport algebra (Zhu and Müller (2023a)).

Another relevant issue is the modeling of noise contamination in metric statistics. The additive noise model commonly employed in Euclidean settings is no longer feasible, as there is no addition operation in metric spaces. However, noise can be modeled by random perturbation maps $\mathcal{P} : \Omega \to \Omega$ that satisfy (Chen and Müller (2022))

$$\omega = \operatorname{argmin}_{x\in\Omega}\mathbb{E}\big[d^2\big(\mathcal{P}(\omega), x\big)\big] \quad \text{for all } \omega \in \Omega.$$

This is the equivalent of the postulate that an additive error e in a Euclidean setting satisfies $\mathbb{E}e = 0$, while $\mathbb{E}(e^2) = \sigma^2$ corresponds to $\mathbb{E}[d^2(\mathcal{P}(\omega), \omega)] = \sigma^2$ in the general case.

In addition to location estimation another important thread in statistics is the estimation of spread, which is essential for uncertainty quantification. Plugging the Fréchet mean into the Fréchet variance function (4) gives the Fréchet variance

$$V_F = \mathbb{E}d^2(X, \mu_\oplus), \qquad \widehat{V}_F = \frac{1}{n}\sum_{i=1}^{n} d^2(X_i, \widehat{\mu}_\oplus), \tag{6}$$

for which under suitable entropy conditions a central limit theorem holds,

$$n^{1/2}(\widehat{V}_F - V_F) \rightsquigarrow N\big(0, \sigma_F^2\big).$$

This can be used to obtain an ANOVA-like test to compare populations of random objects as well as inference for change-points in a sequence of random objects (Dubey and Müller (2019), Dubey and Müller (2020b)).

It is easy to see that when $\Omega = \mathbb{R}$ with the Euclidean metric, the empirical Fréchet variance (6) equals the classical sample variance σ^2. For $X_i \in \mathbb{R}$, it is well known that $\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n-1}(X_i - \bar{X})^2 = \frac{1}{2n(n-1)}\sum_{i,j=1}^{n}(X_i - X_j)^2$ and this also holds in Hilbert spaces, as $\mathbb{E}\langle X - \mathbb{E}X, X - \mathbb{E}X\rangle = \frac{1}{2}\mathbb{E}\langle X - X', X - X'\rangle$ where X' is an independent copy of X. However, the analogous equality does not hold in general metric spaces, and a second option for quantifying spread is then metric variance (Dubey and Müller (2020a)),

$$\operatorname{Var}_\Omega(X) = \frac{1}{2}\mathbb{E}d^2\big(X, X'\big), \quad \text{where } X' \text{ is an independent copy of } X \in \Omega. \tag{7}$$

This notion can also be extended to metric covariance and metric correlation,

$$\operatorname{Cov}_\Omega(X, Y) = \frac{1}{4}\mathbb{E}\big(d^2\big(X, Y'\big) + d^2\big(X', Y\big) - 2d^2(X, Y)\big),$$

$$\rho_\Omega(X, Y) = \frac{\operatorname{Cov}_\Omega(X, Y)}{\sqrt{\operatorname{Cov}_\Omega(X, X)\operatorname{Cov}_\Omega(Y, Y)}}.$$

As already noted, the sample version of $\mathrm{Var}_\Omega(X)$ in the special case $\Omega = \mathbb{R}$ becomes $\widehat{\sigma}^2 = \frac{1}{2n(n-1)} \sum_{i,j=1}^n (X_i - X_j)^2$. An advantage of metric variance/covariance is that these measures do not rely on the potentially arduous task of obtaining the Fréchet mean in a first step. If (Ω, d) is such that $K(x, y) = d^2(x, y)$ is a kernel of negative type (Klebanov (2005)), that is, for all $n > 1$, $x_1, \ldots, x_n \in \Omega$ and $a_1, \ldots, a_n \in \mathbb{R}$ with $\sum_{i=1}^n a_i = 0$ one has $\sum_{i=1}^n \sum_{j=1}^n a_i a_j K(x_i, x_j) \leq 0$, results of Schoenberg (1937) and Schoenberg (1938) imply that metric correlation has the desirable property that $-1 \leq \rho_\Omega(X, Y) \leq 1$. The main distinction between metric correlation and distance correlation, another measure of dependence between paired metric space data (Lyons (2013), Székely and Rizzo (2017)) is that the latter is tailored to measure probabilistic independence rather than to quantify the strength of "positive" or "negative" association, which is the target of metric correlation. The notion of metric covariance is based on pairwise distances between the random objects in a sample for the empirical version and on expected pairwise distances according to the probability measure P in the population version. This motivated us to consider these distances as a basic characteristic of the distributional properties of random objects that are otherwise hard to assess. To quantify this notion, for any fixed $\omega \in \Omega$ the distances to the random objects as determined by the underlying measure P are then of interest. They are captured by the distribution of the distances between ω and any random element taking values in Ω, which then leads to distance profiles indexed by ω. In the following sections, we explore the properties of distance profiles and how optimal transports between the corresponding distributions can be utilized to obtain transport ranks, transport quantiles and inference to compare populations of random objects.

3. Distance profiles, transport ranks and transport quantiles. To introduce and motivate these key notions, we assume that data and random objects of interest are situated in a totally bounded separable metric space (Ω, d). Consider a probability space $(S, \mathcal{S}, \mathbb{P})$, where S is a sample space, $\mathcal{S}$ is a sigma algebra of subsets of S, and $\mathbb{P}$ is a probability measure. A random object X is an Ω-valued random variable, that is, a measurable map $X\colon S \to \Omega$ and P is a Borel probability measure that governs the distribution of X, $X \sim P$, that is, $P(A) = \mathbb{P}(\{s \in S : X(s) \in A\}) =: \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)) =: \mathbb{P}X^{-1}(A)$, for any Borel measurable $A \subseteq \Omega$. For any $\omega \in \Omega$, let F_ω denote the cumulative distribution function (cdf) of the distribution of the distance between ω and a random element X that is distributed according to P. In our notation, we suppress the dependence of F_ω on P and d.

Formally, for any $t \geq 0$, we define the *distance profile* at ω as

$$F_\omega(t) = \mathbb{P}\big(d(\omega, X) \leq t\big), \tag{8}$$

so that F_ω is a one-dimensional distribution that captures the probability mass enclosed by a metric ball in Ω that has center ω and radius t, for all $t \geq 0$. Thus the distance profile at ω is the distribution of the distances that need to be covered to reach other elements of Ω when starting out at ω, as dictated by the distribution P of the random objects X. When $t \to 0$, the distance profile at ω, $F_\omega(t)$, has the form of a small ball probability around ω (Dabo-Niang (2002), Vakhania, Tarieladze and Chobanyan (1987)). An element ω that is centrally located, that is, close to most other elements, will have a distance profile with more mass near 0, in contrast to a distantly located or outlying element whose distance profile will assign mass farther away from 0. If distance profiles have densities, for a centrally located ω the density will have a mode near 0, while the density near 0 will be small for a distantly located ω. Thus $\{F_\omega : \omega \in \Omega\}$ is a family of one-dimensional distributions indexed by Ω that inform about the location of ω relative to $X \sim P$.

The collection of distance profiles $\{F_\omega : \omega \in \Omega\}$ represents the one-dimensional marginals of the stochastic process $\{d(\omega, X)\}_{\omega \in \Omega}$, which is well defined in the sense of the Kolmogorov

existence theorem (see Proposition 1 for details). These simple marginals uniquely characterize the underlying measure P, if (Ω, d) is a metric space such that the kernel $K : \Omega \times \Omega \to \mathbb{R}$ given by $K(\omega, \omega') = d^{\theta}(\omega, \omega')$ for a $\theta > 0$ is of strong negative type (Klebanov (2005), Lyons (2013)). This means that for all Borel probability measures P on Ω and all measurable functions $h : \Omega \to \mathbb{R}$ it holds that $\int_{\Omega}\int_{\Omega} K(\omega, \omega')h(\omega)h(\omega')\,\mathrm{d}P(\omega)\,\mathrm{d}P(\omega') \le 0$ with equality if and only if $h = 0$ P-a.e. Equivalently, for all Borel probability measures P_1, P_2 on Ω one has

$$\begin{aligned}(9)\qquad &\int_{\Omega}\int_{\Omega} K(\omega, \omega')\,\mathrm{d}P_1(\omega)\,\mathrm{d}P_1(\omega') \\ &\quad + \int_{\Omega}\int_{\Omega} K(\omega, \omega')\,\mathrm{d}P_2(\omega)\,\mathrm{d}P_2(\omega') - 2\int_{\Omega}\int_{\Omega} K(\omega, \omega')\,\mathrm{d}P_1(\omega)\,\mathrm{d}P_2(\omega') \le 0,\end{aligned}$$

where equality holds if and only if $P_1 = P_2$; for further discussion, see Section 8. This characterization of the underlying measures motivates the use of distance profiles to obtain information about the complex distribution of the random objects X. Empirical estimates of the distance profiles that will be used for statistical inference are introduced below. A basic result concerning distance profiles is as follows.

PROPOSITION 1. *The stochastic process $\{d(\omega, X)\}_{\omega\in\Omega}$, for which the distance profiles $\{F_\omega : \omega \in \Omega\}$ as defined in (8) are the one-dimensional marginals, is well defined. Suppose that for some $\theta > 0$, (Ω, d^{θ}) is of strong negative type (9) and that P_1, P_2 are two probability measures on Ω. Then $P_1 = P_2$ if and only if $F_{\omega}^{P_1}(u) = F_{\omega}^{P_2}(u)$ for all $\omega \in \Omega$ and $u \ge 0$, where $F_{\omega}^{P_1}$ and $F_{\omega}^{P_2}$ are the distance profiles of ω with respect to P_1 and P_2.*

Consider the distance profile F_X of a random object $X \in \Omega$, $F_X(u) = \mathbb{P}_{X'}(d(X, X') \le u) = \int_S \mathbb{I}(d(X, X'(s)) \le u)\,\mathrm{d}\mathbb{P}(s)$, where X' is an independent copy of X. For each ω, the push-forward map of F_ω to F_X, given by $F_X^{-1}(F_\omega(\cdot))$, determines the optimal transport from the distance profile F_ω to the distance profile F_X. Here and throughout, F^{-1} denotes the quantile function corresponding to a cdf F, $F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \ge u\}$, for $u \in (0, 1)$. We utilize the optimal mass transport map

$$(10)\qquad H_{X,\omega}(u) = F_X^{-1}\big(F_\omega(u)\big) - u, \quad u \ge 0;$$

see, for example, Ambrosio, Gigli and Savaré (2008), to assign a measure of centrality to an element $\omega \in \Omega$ with respect to P. When F_ω is continuous, by a change of variable, the integral

$$(11)\qquad \int H_{X,\omega}(u)\,\mathrm{d}F_\omega(u) = \int_0^1 \{F_X^{-1}(u) - F_\omega^{-1}(u)\}\,\mathrm{d}u$$

provides a summary measure of the mass transfer when transporting F_ω to F_X.

The utility of this notion is that if ω is more centrally located than X with regard to the measure P, we expect the mass transfer to be predominantly from left to right and the magnitude of the integral in (11) to reflect the outlyingness differential between ω and a random object X, $X \sim P$. For example, for a distribution P that is symmetric around a central point ω_0 and assigns less mass when moving away from ω_0, we expect that the integral (11) with $\omega = \omega_0$ is relatively large and the magnitude of the integral (11) is decreasing as the distance from ω_0 increases. This motivates to take the expected value of the integral in (11) to quantify the degree of centrality or outlyingness of an element ω.

An illustration is in Figure 1 for the simple case where X is a bivariate Gaussian random variable with mean zero and covariance $\mathrm{diag}(2, 1)$. For the points $x \in \{(0, 0), (2, 0), (4, 0), (6, 0)\}$ and $\omega = (2, 2)$, their corresponding distance profiles are depicted as densities f_x and f_ω in the left panel, where the distances from x to the rest of the data

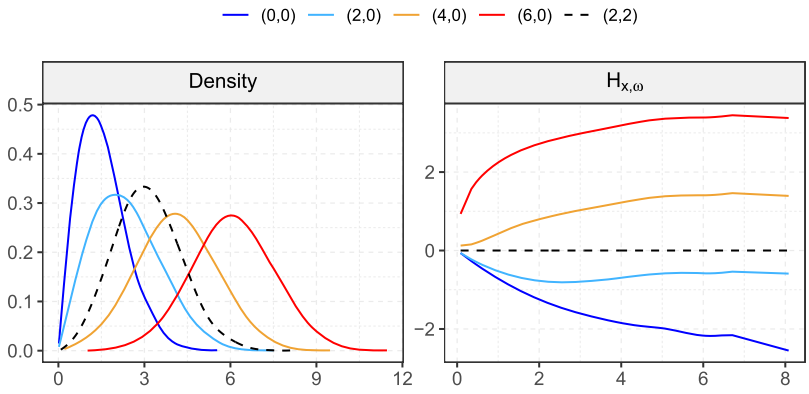

FIG. 1. *Left*: *Distance profiles, represented by the corresponding densities, at five points as indicated, with respect to a bivariate Gaussian distribution with mean zero and covariance* diag(2, 1). *Right*: *Transport maps subtracted by identity* $H_{x,\omega}$ *as per* (10) *for* $x \in \{(0,0), (2,0), (4,0), (6,0)\}$ *and* $\omega = (2,2)$, *where negative* (*positive*) *values indicate transport to the right* (*left*).

are seen to increase as x moves away from the origin, which is exactly what one expects. For $x \in \{(0,0), (2,0), (4,0), (6,0)\}$, the transport maps $H_{x,\omega}$ as per (10) that move mass from F_ω to F_x for the fixed element $\omega = (2,2)$ are in the right panel. For $\omega = (2,2)$, mass moves to the left when transporting F_ω to F_x for $x \in \{(0,0), (2,0)\}$, which are closer to the origin, and moves to the right for $x \in \{(4,0), (6,0)\}$, which are farther away from the origin. Another example based on the U.S. electricity generation compositional data in Section 7.4 is shown in Figure 2. Transporting mass from the distance profile of $\omega =$ New Jersey (NJ) to the profiles x of Maryland (MD), Massachusetts (MA), Louisiana (LA) to Rhode Island (RI), one moves from the profile of a point in the middle of the ternary plot toward the profiles of points closer to the boundary of the compositional space. The mass transport moves mass mostly to the left when transporting F_ω to F_x for $x = \{\text{MD}, \text{MA}\}$ and unambiguously to the right for $x = \text{RI}$.

This motivates the notion of *transport ranks* to measure centrality of an element $\omega \in \Omega$ with respect to P as the expit of the expected integrated mass transfer when transporting F_ω

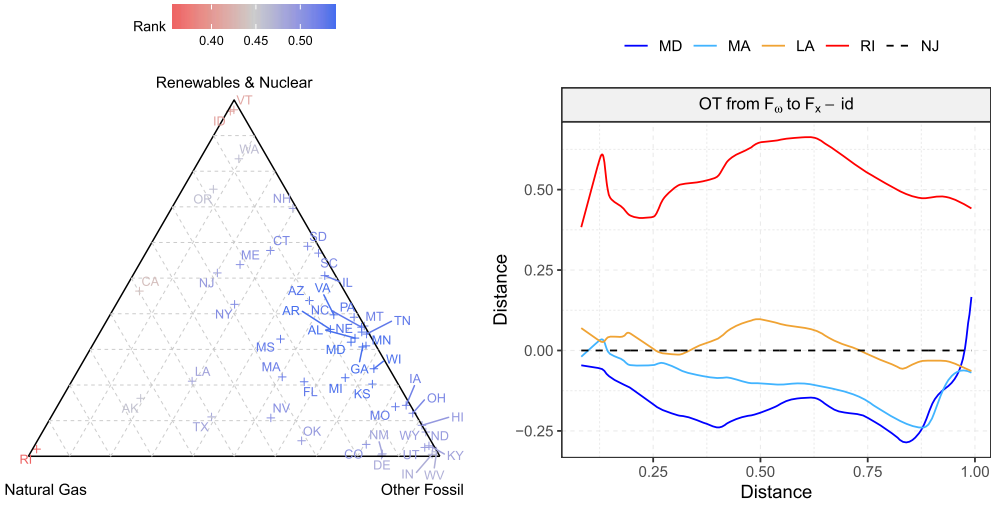

FIG. 2. *Left*: *Ternary plot of compositions of electricity generation in the year* 2000 *for the* 50 *states in the U.S., where the points are colored according to their transport ranks. Right*: *Mass transport maps* $H_{x,\omega}$ *as per* (10) *for the transports from* $\omega = NJ$ *to* $x \in \{MD, MA, LA, RI\}$.

to F_X, where $P = \mathbb{P}X^{-1}$. We use the expit function $\operatorname{expit}(x) = e^x/(1+e^x)$ in the definition of these ranks to ensure that the proposed ranks are scaled to lie in $(0, 1)$; any strictly monotone invertible function from $\mathbb{R}$ to $[0, 1]$ can be used for this purpose. Formally,

$$R_\omega = \operatorname{expit}\left[\mathbb{E}\left\{\int_0^1 \left[F_X^{-1}(u) - F_\omega^{-1}(u)\right] \mathrm{d}u\right\}\right]. \tag{12}$$

The transport rank of ω quantifies the aggregated preference of ω with respect to the data cloud. The greater the transport rank of ω is, the more centered ω is relative to the sample elements. Equipped with an ordering of the elements of Ω by means of their transport ranks, we define the *transport median* set $\mathcal{M}_\oplus$ of P as the collection of points in the support $\Omega_P \subset \Omega$ of P, which have maximal transport rank and are therefore most central,

$$\mathcal{M}_\oplus = \operatorname*{argmax}_{\omega \in \Omega_P} R_\omega. \tag{13}$$

The distance profiles of the data objects together with the transport ranks and the transport median set are the key ingredients of the proposed toolkit to quantify centrality. These devices lend themselves to devise distance profile based methods for cluster analysis, classification and outlier detection, all of which are challenging when one deals with random objects. The set of maximizers of R_ω in Ω_P constitutes the transport median set defined in (13). Observing that the function R_ω is uniformly continuous in ω by Lemma S.3 in the Supplementary Material, the transport median set is guaranteed to be nonempty whenever Ω_P is compact. If Ω is a length space that is complete and locally compact, the Hopf–Rinow theorem (Chavel (2006)) implies that if Ω_P is any bounded closed subset of Ω, it is guaranteed to be compact.

Once a center-outward ordering of the elements of Ω has been established through their transport ranks, these ranks can be utilized in numerous ways. One application is to define level sets of the form $L_\alpha = \{\omega \in \Omega : R_\omega = \alpha\}$ and nested superlevel sets $L_\alpha^+ = \{\omega \in \Omega : R_\omega \geq \alpha\}$. By definition, $L_{\alpha_1}^+ \subseteq L_{\alpha_2}^+$ whenever $\alpha_1 \geq \alpha_2$. Due to the continuity of R_ω (see Lemma S.3 in the Supplementary Material), the sets L_α and L_α^+ are closed. Moreover when (Ω, d) is a bounded, complete and locally compact length space, again by the Hopf–Rinow theorem L_α and L_α^+ are compact as well. Superlevel sets L_α^+ can be used to define *transport quantile sets*. These can be viewed as a generalization of univariate quantiles to general random objects. Specifically, a ζ-level transport quantile set can be defined as the intersection of level sets L_α such that $\mathbb{P}(X \in L_\alpha^+) \geq \zeta$, for $\zeta \in (0, 1)$. Complements of superlevel sets can be used to identify potential outliers by highlighting observations with low transport ranks. Data trimming can be achieved by excluding points that have transport ranks lower than a suitably chosen threshold α_0; one then might consider maximizers of transport ranks over trimmed versions of Ω_P to obtain trimmed analogues of the transport median set $\mathcal{M}_\oplus$ and also trimmed Fréchet means.

4. Properties of distance profiles and transport ranks. We discuss here some desirable properties of distance profiles, transport ranks and the transport median set that are appropriately modified versions of analogous properties of classical ranks.

Lipschitz continuity of transport ranks. By Lemma S.2 in Section S.3 in the Supplementary Material, the distance profiles $F_\omega(\cdot)$ and the associated quantile function representations $F_\omega^{-1}(\cdot)$ are uniformly Lipschitz in ω provided that the distance profiles have uniformly upper bounded densities with respect to the Lebesgue measure. This means that F_{ω_1} and F_{ω_2} are uniformly close to each other as long as ω_1 and ω_2 are close, and the distance between F_{ω_1} and F_{ω_2} is upper bounded by a constant factor of $d(\omega_1, \omega_2)$. Moreover transport ranks R_ω are uniformly Lipschitz in ω, see Lemma S.3 in the Supplementary Material.

Invariance of transport ranks. Let $(\tilde{\Omega}, \tilde{d})$ be a metric space. A map $h\colon \Omega \to \tilde{\Omega}$ is isometric if $d(\omega_1, \omega_2) = \tilde{d}(h(\omega_1), h(\omega_2))$ for all $\omega_1, \omega_2 \in \Omega$. Theorem 4.1(a) establishes the invariance of distance profiles, and thereby of transport ranks, under isometric transformations. In Euclidean spaces, this ensures that the distance profiles are invariant under orthogonal transformations such as rotations.

Transport modes and center-outward decay of transport ranks. Consider a situation where the distribution P of X concentrates around a point $\omega_\oplus \in \Omega$. Specifically, if there exists an element $\omega_\oplus \in \Omega$ such that

$$F_{\omega_\oplus}(u) \geq F_\omega(u) \tag{14}$$

for any $\omega \in \Omega$ and any $u \geq 0$, we refer to $\omega_\oplus$ as an Ω-valued *transport mode* of P. Condition (14) states that a d-ball of radius u around $\omega_\oplus$ contains more mass under P than a similar ball around any other point in Ω. According to Theorem 4.1(b), if P has a transport mode, then the transport rank of the transport mode cannot be smaller than that of any other $\omega \in \Omega$ and therefore, a transport mode is always contained in the transport median set. For distributions that concentrate around their unique Fréchet mean (Fréchet (1948)), the Fréchet mean is the transport mode, and hence is contained in the transport median set; for the special case of networks, see Lunagómez, Olhede and Wolfe (2021). Theorem 4.1(c) provides a characterization of the radial ordering induced by the transport rank for the special case where the data distribution on Ω has a transport mode $\omega_\oplus$, by considering curves of the form $\gamma\colon [0, 1] \to \Omega$ that originate from $\omega_\oplus$, that is, $\gamma(0) = \omega_\oplus$. According to Theorem 4.1(c), transport ranks are nonincreasing along curves originating from a transport mode $\omega_\oplus$, whenever P is such that the distance profiles decay systematically along the curve $\gamma(t)$ as t is increasing.

Characterization of the probability measure P through transport ranks. Theorem 4.1(d) shows that when (Ω, d) is of strong negative type, the comprehensive set of all transport ranks $\{R_\omega\}_{\omega\in\Omega}$ uniquely characterizes the underlying measure.

THEOREM 4.1. *For a separable metric space (Ω, d) the distance profiles F_ω and the transport ranks R_ω satisfy the following properties*:

(a) *Let $h\colon \Omega \to \tilde{\Omega}$ be a bijective isometric measurable map between (Ω, d) and $(\tilde{\Omega}, \tilde{d})$ and $P_h(\cdot) = P(h^{-1}(\cdot))$ the push-forward measure on $\tilde{\Omega}$. Then $F^{P_h}_{h(\omega)}(u) = F^P_\omega(u)$ for all $u \in \mathbb{R}$, hence $R^{P_h}_{h(\omega)} = R^P_\omega$, where $F^P_\omega(u) = \mathbb{P}(d(\omega, X) \leq u)$ and X is a Ω-valued random element such that $P = \mathbb{P}X^{-1}$, $F^{P_h}_{h(\omega)}(u) = \mathbb{P}(\tilde{d}(h(\omega), h(X)) \leq u)$, R^P_ω is the transport rank of ω with respect to P and $R^{P_h}_{h(\omega)}$ is the transport rank of $h(\omega)$ with respect to P_h.*

(b) *If $\omega_\oplus$ is a transport mode of P as per (14), $R_{\omega_\oplus} \geq 1/2$. Moreover $R_{\omega_\oplus} \geq R_\omega$ for any $\omega \in \Omega$ and $\omega_\oplus \in \mathcal{M}_\oplus$.*

(c) *Suppose $\omega_\oplus$ is a transport mode of P. Let $\gamma\colon [0, 1] \to \Omega$ be curve in (Ω, d) such that $\gamma(0) = \omega_\oplus$ and $F_{\gamma(s)}(u) \geq F_{\gamma(t)}(u)$ for all $u \in \mathbb{R}$ and $0 \leq s < t \leq 1$. Then $R_{\gamma(s)}(u) \geq R_{\gamma(t)}(u)$ for all $u \in \mathbb{R}$ and $0 \leq s < t \leq 1$.*

(d) *Suppose the metric space (Ω, d) is of strong negative type (9) and P_1, P_2 are two probability measures on the space. Then $P_1 = P_2$ if and only if $R^{P_1}_\omega = R^{P_2}_\omega$ for all $\omega \in \Omega$, where $R^{P_1}_\omega$ and $R^{P_2}_\omega$ are the transport ranks of ω with respect to P_1 and P_2.*

5. Estimation and large sample properties. While so far we have introduced the notions of distance profiles, transport ranks and transport median sets at the population level, in practice one needs to estimate these quantities from a data sample of random objects $\{X_i\}_{i=1}^n$

consisting of n independent realizations of X. To obtain the distance profiles F_ω, $\omega \in \Omega_P$ from a sample, we use empirical estimates

$$\widehat{F}_\omega(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\big(d(\omega, X_i) \le t\big), \quad t \ge 0, \tag{15}$$

where $\mathbb{I}(A)$ is the indicator function for an event A.

Replacing expectations with empirical means and using estimated distance profiles $\widehat{F}_{X_i}$ given by $\widehat{F}_{X_i}(t) = \frac{1}{n-1}\sum_{1\le j\le n, j\ne i} \mathbb{I}(d(X_j, X_i) \le t)$ for $t \ge 0$ as surrogates of F_{X_i}, we obtain estimates for the transport rank of $\omega \in \Omega$ defined in (12) as

$$\widehat{R}_\omega = \operatorname{expit}\left[\frac{1}{n}\sum_{i=1}^{n}\left\{\int_0^1 [\widehat{F}_{X_i}^{-1}(u) - \widehat{F}_\omega^{-1}(u)]\,\mathrm{d}u\right\}\right]. \tag{16}$$

The term $\int_0^1 [\widehat{F}_{X_i}^{-1}(u) - \widehat{F}_\omega^{-1}(u)]\,\mathrm{d}u$ provides a comparison between the outlyingness of ω and that of X_i; mass movement predominantly to the right (left) indicates that ω is more central (outlying) compared to X_i, respectively. Finally, we define the estimated transport median set

$$\widehat{\mathcal{M}}_\oplus = \operatorname*{argmax}_{\omega\in\{X_1, X_2, \ldots, X_n\}} \widehat{R}_\omega. \tag{17}$$

To obtain asymptotic properties of these estimators and convergence toward their population targets, we require the following assumptions.

ASSUMPTION 1. Let $N(\varepsilon, \Omega, d)$ be the covering number of the space Ω with balls of radius ε and $\log N(\varepsilon, \Omega, d)$ the corresponding metric entropy. Then

$$\varepsilon \log N(\varepsilon, \Omega, d) \to 0 \quad \text{as } \varepsilon \to 0. \tag{18}$$

ASSUMPTION 2. For every $\omega \in \Omega$, F_ω is absolutely continuous with continuous density f_ω. For $\underline{\Delta}_\omega = \inf_{t\in\operatorname{supp}(f_\omega)} f_\omega(t)$ and $\overline{\Delta}_\omega = \sup_{t\in\mathbb{R}} f_\omega(t)$, $\underline{\Delta}_\omega > 0$ for each $\omega \in \Omega$ and there exists $\overline{\Delta} > 0$ such that $\sup_{\omega\in\Omega} \overline{\Delta}_\omega \le \overline{\Delta}$.

Assumptions 1 and 2 are required for Theorem 5.1, which provides the uniform convergence of $\widehat{F}_\omega$ to F_ω. The entropy condition Assumption 1 serves to overcome the dependence between the summands in the estimator of the transport rank and to establish uniform convergence to the population transport ranks. Assumption 2 is a smoothness condition of the probability measure P, which is widely satisfied. Examples include absolutely continuous distributions with compact support in a Euclidean space or probability distributions on a Riemannian manifold, where, after applying Riemannian log maps, the transformed distributions on tangent spaces are absolutely continuous with compact support.

For any $t \ge 0$, $F_\omega(t) = \mathbb{E}(\widehat{F}_\omega(t))$. For functions $y_{\omega,t}\colon \Omega \to \mathbb{R}$ with $y_{\omega,t}(x) = \mathbb{I}\{d(\omega, x) \le t\}$ and the function class $\mathcal{F} = \{y_{\omega,t} : \omega \in \Omega, t \in \mathbb{R}\}$, the following result establishes that under Assumptions 1 and 2 the function class $\mathcal{F}$ is P-Donsker.

THEOREM 5.1. *Under Assumptions 1 and 2, $\{\sqrt{n}(\widehat{F}_\omega(t) - F_\omega(t)) : \omega \in \Omega, t \in \mathbb{R}\}$ converges weakly to a zero-mean Gaussian process $\mathbb{G}_P$ with covariance given by*

$$\mathcal{C}_{(\omega_1,t_1),(\omega_2,t_2)} = \operatorname{Cov}\big(y_{\omega_1,t_1}(X), y_{\omega_2,t_2}(X)\big)$$

for $\omega_1, \omega_2 \in \Omega$ and $t_1, t_2 \in \mathbb{R}$.

Assumption 1 is a restriction on the complexity of the metric space (Ω, d). It is satisfied for a broad class of spaces. In particular, any space (Ω, d) such that $\log N(\varepsilon, \Omega, d) = O(\frac{1}{\varepsilon^\alpha})$ for some $\alpha < 1$ satisfies Assumption 1. This is true for any (Ω, d), which can be represented as a subset of elements in a finite-dimensional Euclidean space, for example, the space of graph Laplacians or network adjacency matrices with a fixed number of nodes (Ginestet et al. (2017), Kolaczyk et al. (2020)), symmetric positive definite matrices of a fixed size (Dryden, Koloydenko and Zhou (2009)), simplex valued objects in a fixed dimension (Jeon and Park (2020)) and the space of phylogenetic trees with the same number of tips (Billera, Holmes and Vogtmann (2001), Kim, Rosenberg and Palacios (2020)). It holds that $\log N(\varepsilon, \Omega, d) = O(\varepsilon^{-\alpha})$ for any $\alpha < 1$ when Ω is a VC-class of sets or a VC-class of functions (van der Vaart and Wellner (1996), Theorems 2.6.4 and 2.6.7). Assumption 1 also holds for p-dimensional smooth function classes $C_1^\alpha(\mathcal{X})$ (van der Vaart and Wellner (1996), p. 155) on bounded convex sets $\mathcal{X}$ in $\mathbb{R}^p$ equipped with the $\|\cdot\|_\infty$-norm (van der Vaart and Wellner (1996), Theorem 2.7.1) or the $\|\cdot\|_{r,Q}$-norm for any probability measure Q on $\mathbb{R}^p$ (van der Vaart and Wellner (1996), Corollary 2.7.2), if $\alpha \geq p + 1$.

Of particular interest for many applications is the case when Ω is the space of one-dimensional distributions on some compact interval $I \subset \mathbb{R}$ with the 2-Wasserstein metric $d = d_W$ defined in (1) (Petersen and Müller (2019)). If Ω is represented using the quantile functions of the distributions then, without any further assumptions, $\log N(\varepsilon, \Omega, d_W)$ is upper and lower bounded by a factor of $1/\varepsilon$ (Blei, Gao and Li (2007), Proposition 2.1) and does not meet the criterion in Assumption 1. However, if we assume that the distributions in Ω are absolutely continuous with respect to the Lebesgue measure on I with smooth densities uniformly taking values in some interval $[l_0, u_0]$, $0 < l_0 < u_0 < \infty$, then Ω equipped with d_W satisfies Assumption 1. To see this, observe that with the above characterization of Ω the quantile functions corresponding to the distributions in Ω have smooth derivatives that are uniformly bounded. With $\mathcal{Q}_{\text{deriv}}$ denoting the space of the uniformly bounded derivatives of the quantile functions in Ω, $\log N(\varepsilon, \mathcal{Q}_{\text{deriv}}, \|\cdot\|_1) = O(\varepsilon^{-1})$, where $\|\cdot\|_1$ is the L_1 norm under the Lebesgue measure on I (van der Vaart and Wellner (1996), Corollary 2.7.2). Using Lemma 1 in Gao and Wellner (2009), with $\mathcal{F} \equiv \mathcal{Q}_{\text{deriv}}$, $\mathcal{G} \equiv \Omega$, $\alpha(x) = x$ and $\phi(\varepsilon) = K/\varepsilon$ for some constant K, $\log N(\varepsilon, \Omega, d_W) = O(\varepsilon^{-1/2})$, which meets the requirement of Assumption 1. If Ω is the space of p-dimensional distributions on a compact convex set $I \subset \mathbb{R}^p$, represented using their distribution functions endowed with the L_r metric with respect to the Lebesgue measure on I, then Assumption 1 is satisfied if $\Omega \subset C_1^\alpha(I)$ for $\alpha \geq p + 1$.

Next, we discuss the asymptotic convergence of the estimates $\widehat{R}_\omega$ of transport ranks. Theorem 5.2 establishes a $\sqrt{n}$-rate of convergence uniformly in ω.

THEOREM 5.2. *Under Assumptions 1 and 2,*

$$\sqrt{n} \sup_{\omega \in \Omega} |\widehat{R}_\omega - R_\omega| = O_{\mathbb{P}}(1).$$

To conclude this section, we consider the convergence of the estimated transport median set $\widehat{\mathcal{M}}_\oplus$ to $\mathcal{M}_\oplus$ in the Hausdorff metric,

$$\rho_H(\widehat{\mathcal{M}}_\oplus, \mathcal{M}_\oplus) = \max\Big(\sup_{\omega \in \widehat{\mathcal{M}}_\oplus} d(\omega, \mathcal{M}_\oplus), \sup_{\omega \in \mathcal{M}_\oplus} d(\omega, \widehat{\mathcal{M}}_\oplus) \Big), \tag{19}$$

where for any $\omega \in \Omega$ and any subset $A \subset \Omega$, $d(\omega, A) = \inf_{s \in A} d(\omega, s)$. We derive uniform Lipschitz continuity of transport ranks R_ω in ω (Lemma S.3 in the Supplementary Material) and require the following additional assumption.

ASSUMPTION 3. For some $\eta' > 0$, for any $0 < \varepsilon < \eta'$,

$$\alpha(\varepsilon) = \inf_{\tilde{\omega} \in \mathcal{M}_\oplus} \inf_{d(\omega, \tilde{\omega}) > \varepsilon} |R_\omega - R_{\tilde{\omega}}| > 0.$$

Assumption 3 deals with the identifiability of transport medians and stipulates that the transport median set is a union of single point sets, which are separated from each other by a minimum fixed distance. In particular, unimodal probability measures satisfy Assumption 3. This assumption is needed to derive our next result on the consistency of the estimated transport median set for the true transport median set in the Hausdorff metric.

THEOREM 5.3. *Assume that the distribution P is such that $\mathcal{M}_\oplus$ is nonempty. Under Assumptions 1–3,*

$$\rho_H(\widehat{\mathcal{M}}_\oplus, \mathcal{M}_\oplus) = o_{\mathbb{P}}(1).$$

6. Two-sample inference with distance profiles.

6.1. *Construction of a two-sample test.* Assume that $X_1, X_2, \ldots, X_n$ is a sample of random objects taking values in Ω, generated according to a Borel probability measure P_1 on Ω, and that $Y_1, Y_2, \ldots, Y_m$ is another sample of Ω-valued random objects generated analogously according to a Borel probability measure P_2. Two-sample testing in this setting concerns the null (20) and alternative (21) hypotheses

$$H_0 : P_1 = P_2, \tag{20}$$

$$H_1 : P_1 \neq P_2. \tag{21}$$

Nonparametric two-sample tests have been studied extensively in many settings. To extend this classical problem to object data poses new challenges. Existing methods that are based on distances, such as the graph based tests (Chen and Friedman (2017)) and the energy test (Székely and Rizzo (2004)), either require tuning parameters for their practical implementation or lack theoretical guarantees on the power of the test, particularly when using permutation cut-offs for type I error control. We propose here a two-sample test based on the distance profiles of the observations. The proposed test is tuning parameter-free, has rigorous asymptotic type I error control under H_0 (20) and is guaranteed to be powerful against contiguous alternatives for sufficiently large sample sizes. While it was presented at the Rietz Lecture and derived independently, our test statistic is similar in spirit to a test proposed in Wang et al. (2023). We note that our results are derived under weaker assumptions and provide, in addition to consistency under the null hypothesis, power guarantees under contiguous alternatives, as well as theoretical guarantees for the corresponding permutation tests.

We require some notation. For $\omega \in \Omega$, the distance profiles of ω with respect to $X \sim P_1$ and $Y \sim P_2$, respectively, are given by $F_\omega^X(\cdot)$ and $F_\omega^Y(\cdot)$, where for $u \in \mathbb{R}$,

$$F_\omega^X(u) = \mathbb{P}\big(d(x, X) \leq u\big) \quad \text{and} \quad F_\omega^Y(u) = \mathbb{P}\big(d(x, Y) \leq u\big). \tag{22}$$

Let $\widehat{F}_{X_1}^X(\cdot), \widehat{F}_{X_2}^X(\cdot), \ldots, \widehat{F}_{X_n}^X(\cdot)$ be the estimated *in-sample* distance profiles of $X_1, \ldots, X_n$, respectively, with respect to the observations from P_1, that is,

$$\widehat{F}_{X_i}^X(u) = \frac{1}{n-1} \sum_{j \neq i} \mathbb{I}\big(d(X_i, X_j) \leq u\big).$$

Then we obtain the *out-of-sample* distance profiles of $X_1, \ldots, X_n$, respectively, with respect to the observations from P_2, through $\widehat{F}_{X_1}^Y(\cdot), \widehat{F}_{X_2}^Y(\cdot), \ldots, \widehat{F}_{X_n}^Y(\cdot)$, where

$$\widehat{F}_{X_i}^Y(u) = \frac{1}{m} \sum_{j=1}^{m} \mathbb{I}\big(d(X_i, Y_j) \leq u\big).$$

Similarly, we estimate the in-sample and the out-of-sample distance profiles of $Y_1, \ldots, Y_m$ with respect to the observations from P_2 and P_1, respectively, given by $\widehat{F}^Y_{Y_1}(\cdot)$, $\widehat{F}^Y_{Y_2}(\cdot), \ldots, \widehat{F}^Y_{Y_m}(\cdot)$ and $\widehat{F}^X_{Y_1}(\cdot)$, $\widehat{F}^X_{Y_2}(\cdot), \ldots, \widehat{F}^X_{Y_m}(\cdot)$, respectively, where for $u \geq 0$,

$$\widehat{F}^Y_{Y_j}(u) = \frac{1}{m-1} \sum_{j \neq i} \mathbb{I}\big(d(Y_j, Y_i) \leq u\big)$$

and

$$\widehat{F}^X_{Y_j}(u) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\big(d(Y_j, X_i) \leq u\big).$$

With T^X_{nm} and T^Y_{nm} defined as

$$T^X_{nm}(X, Y) = \frac{1}{n} \sum_{i=1}^{n} \int \big\{\widehat{F}^X_{X_i}(u) - \widehat{F}^Y_{X_i}(u)\big\}^2 \, \mathrm{d}u$$

and

$$T^Y_{nm}(X, Y) = \frac{1}{m} \sum_{i=1}^{m} \int \big\{\widehat{F}^X_{Y_i}(u) - \widehat{F}^Y_{Y_i}(u)\big\}^2 \, \mathrm{d}u,$$

the proposed test statistic $T_{nm}(X, Y)$ is

$$T_{nm}(X, Y) = \frac{nm}{n+m} \big\{T^X_{nm} + T^Y_{nm}\big\}. \tag{23}$$

To enhance flexibility, we also consider a generalized weighted version of the test statistic, where for each observation X_i or Y_i, we allow for data adaptive weight profiles $\widehat{w}_{X_i}(\cdot)$ and $\widehat{w}_{Y_i}(\cdot)$ that can be tuned appropriately to enhance the detection capacity of the test statistic. This leads to weighted versions of T^X_{nm} and T^Y_{nm},

$$T^{X,w}_{nm}(X, Y) = \frac{1}{n} \sum_{i=1}^{n} \int \widehat{w}_{X_i}(u) \big\{\widehat{F}^X_{X_i}(u) - \widehat{F}^Y_{X_i}(u)\big\}^2 \, \mathrm{d}u$$

$$\text{and} \quad T^{Y,w}_{nm}(X, Y) = \frac{1}{m} \sum_{i=1}^{m} \int \widehat{w}_{Y_i}(u) \big\{\widehat{F}^X_{Y_i}(u) - \widehat{F}^Y_{Y_i}(u)\big\}^2 \, \mathrm{d}u$$

and the weighted test statistic

$$T^w_{nm}(X, Y) = \frac{nm}{n+m} \big\{T^{X,w}_{nm} + T^{Y,w}_{nm}\big\}. \tag{24}$$

Note that the test statistic in equation (23) is a version of the generalized test statistic in equation (24) with $\widehat{w}_{X_i}(\cdot) = \widehat{w}_{Y_i}(\cdot) \equiv 1$. Hereafter, we suppress the dependence of T^X_{nm}, T^Y_{nm}, T_{nm}, $T^{X,w}_{nm}$, $T^{Y,w}_{nm}$ and T^w_{nm} on (X, Y) as this will be clear from the context.

Suppose that for each $x \in \Omega$, there exists a population limit of the estimated data adaptive weight profile $\widehat{w}_x(\cdot)$ given by $w_x(\cdot)$ such that

$$\sup_{x \in \Omega} \sup_{u} \big|\widehat{w}_x(u) - w_x(u)\big| = o_{\mathbb{P}}(1). \tag{25}$$

The weight profile dependent quantities

$$\begin{aligned} D^w_{XY}(P_1, P_2) = {} & \mathbb{E}\bigg\{\int w_{X'}(u)\big(F^X_{X'}(u) - F^Y_{X'}(u)\big)^2 \, \mathrm{d}u\bigg\} \\ & + \mathbb{E}\bigg\{\int w_{Y'}(u)\big(F^X_{Y'}(u) - F^Y_{Y'}(u)\big)^2 \, \mathrm{d}u\bigg\}, \end{aligned} \tag{26}$$

where $F_\omega^X(\cdot)$ and $F_\omega^Y(\cdot)$ are as defined in equation (22) and $X' \sim P_1$ and $Y' \sim P_2$, capture the population version of the proposed test statistic (24). First, observe that under H_0, for any $w_x(\cdot), x \in \Omega$, it holds that $D_{XY}^w(P_1, P_2) = 0$. Next, we show that under mild conditions on the weight profiles, (Ω, d), P_1 and P_2, $D_{XY}^w(P_1, P_2) = 0$ if and only if $P_1 = P_2$.

Let $\mathcal{P}_{(\Omega,d)}$ denote the class of all Borel probability measures on (Ω, d) that are uniquely determined by the measure of all open balls or equivalently by the set of distance profiles, that is, for $Q_1, Q_2 \in \mathcal{P}_{(\Omega,d)}$, $Q_1 = Q_2$ if and only if $F_\omega^{Q_1}(u) = F_\omega^{Q_2}(u)$ for all $\omega \in \Omega$ and $u \geq 0$. In fact, $\mathcal{P}_{(\Omega,d)}$ corresponds to the set of all Borel probability measures on (Ω, d) when the metric d is such that d^θ is of strong negative type (9) for some $\theta > 0$ (see Proposition 1). If $w_\omega(u) > 0$ for any $\omega \in \Omega$ and for any $u \geq 0$, then $D_{XY}^w = 0$ implies that $F_\omega^X(u) = F_\omega^Y(u)$ for almost any $u \geq 0$ and for any ω in the union of the supports of P_1 and P_2. Hence, if Ω is contained in the union of the supports of P_1 and P_2, then $D_{XY}^w(P_1, P_2) = 0$ implies that $P_1 = P_2$ whenever $P_1, P_2 \in \mathcal{P}_{(\Omega,d)}$. In the following, we will suppress (P_1, P_2) in the notation $D_{XY}^w(P_1, P_2)$.

We will use D_{XY}^w in Section 6.2 to evaluate the power performance of the test by constructing a sequence of contiguous alternatives that approach the null hypothesis (20). To obtain the asymptotic power of the test, we work with q_α, the asymptotic critical value for rejecting H_0, where $q_\alpha = \inf\{t : \Gamma_L(t) \geq 1 - \alpha\}$ and $\Gamma_L(\cdot)$ is the cumulative distribution function of the asymptotic null distribution corresponding to the law of L (see Theorem 6.1). Since L is an infinite mixture of chi-squares with mixing weights depending on the data distribution under H_0, we estimate q_α using a random permutation scheme in practice, as follows.

Let Π_{nm} denote the collection of all $(n + m)!$ permutations of $\{1, 2, \ldots, n + m\}$ and Π a random variable that follows a uniform distribution on Π_{nm} and is independent of the observations $X_1, \ldots, X_n$ and $Y_1, \ldots Y_m$. Let $V_1, V_2, \ldots, V_{n+m}$ denote the pooled sample where $V_i = X_i$ if $i \leq n$ and $V_i = Y_{i-n}$ if $i \geq n + 1$. Let $\Pi_1, \Pi_2, \ldots \Pi_K$ denote i.i.d. replicates of Π. Each $\Pi_j = (\Pi_j(1), \ldots, \Pi_j(n + m))$ is a random permutation of $\{1, 2, \ldots, n + m\}$ and when applied to the data yields $V_{\Pi_j} = \{V_{\Pi_j(1)}, V_{\Pi_j(2)}, \ldots, V_{\Pi_j(n+m)}\}$, $j = 1, \ldots, K$, which constitute a collection of randomly permuted pooled data. For each $j = 1, \ldots, K$, split the data V_{Π_j} into $X_{\Pi_j} = \{V_{\Pi_j(1)}, V_{\Pi_j(2)}, \ldots, V_{\Pi_j(n)}\}$ and $Y_{\Pi_j} = \{V_{\Pi_j(n+1)}, V_{\Pi_j(n+2)}, \ldots, V_{\Pi_j(n+m)}\}$. With X_{Π_j} and Y_{Π_j} being the proxies for the two samples of sizes n and m, respectively, evaluate the test statistic replicates $T_{\Pi_j} = T_{nm}^w(X_{\Pi_j}, Y_{\Pi_j})$. Define $\widehat{\Gamma}_{nm}(\cdot)$ as

$$\widehat{\Gamma}_{nm}(t) = \frac{1}{K} \sum_{j=1}^{K} \mathbb{I}(T_{\Pi_j} \leq t), \tag{27}$$

which approximates the randomization distribution of T_{nm}^w using the random permutations $\Pi_1, \Pi_2, \ldots, \Pi_K$. Then a natural estimate of q_α is

$$\widehat{q}_\alpha = \inf\{t : \widehat{\Gamma}_{nm}(t) \geq 1 - \alpha\}. \tag{28}$$

6.2. *Theoretical guarantees for type I error control and the asymptotic power of the test.* To establish theoretical guarantees of the proposed test, in particular the limiting distribution of the test statistic under H_0 (20) and the consistency of the test under the alternative (21), we require additional assumptions, including a modified version of Assumption 2, listed below as Assumption 5. Assumption 4 requires regularity conditions on the data adaptive weight profiles to ensure that they have a well behaved asymptotic limit. Assumption 6 is needed so that none of the group sizes is asymptotically negligible.

ASSUMPTION 4. For each $x \in \Omega$, there exists a population limit of the estimated weight profiles such that (25) is satisfied; there exists $C_w > 0$ such that $\sup_{x\in\Omega} \sup_u |w_x(u)| \leq C_w$; for some $L_w > 0$ it holds that $\sup_{x\in\Omega} |w_x(u) - w_x(v)| \leq L_w|u - v|$.

ASSUMPTION 5. For each $x \in \Omega$, $X \sim P_1$ and $Y \sim P_2$, $F_x^X(t) = \mathbb{P}(d(x, X) \le t)$ and $F_x^Y(t) = \mathbb{P}(d(x, Y) \le t)$ are absolutely continuous, with densities $f_x^X(t)$ and $f_x^Y(t)$, respectively, that satisfy $\inf_{t \in \operatorname{supp}(f_x^X)} f_x^X(t) > 0$, $\inf_{t \in \operatorname{supp}(f_x^Y)} f_x^Y(t) > 0$. There exist $L_X, L_Y > 0$ such that $\sup_{x \in \Omega} \sup_{t \in \mathbb{R}} |f_x^X(t)| \le L_X$ and $\sup_{x \in \Omega} \sup_{t \in \mathbb{R}} |f_x^Y(t)| \le L_Y$.

ASSUMPTION 6. There exists $0 < c < 1$ such that sample sizes n and m satisfy $n/(n+m) \to c$ as $n, m \to \infty$.

Theorem 6.1 provides the framework for asymptotic type I error control of the test. The asymptotic distribution of the test statistic T_{nm} (23), which we will illustrate later in the simulations, can be directly derived from Theorem 6.1 by plugging in $w_{X_i}(\cdot) = w_{Y_i}(\cdot) \equiv 1$ as this constant weight profile satisfies Assumption 4 trivially.

THEOREM 6.1. *Under H_0 (20) and Assumptions 1, 4, 5 and 6, T_{nm}^w converges in distribution to the law of a random variable $L = 2\sum_{j=1}^{\infty} Z_j^2 \mathbb{E}_V(\lambda_j^V)$, where $Z_1, Z_2, \ldots$ is a sequence of i.i.d. $N(0,1)$ random variables, $V \sim P$ where $P = P_1 = P_2$ under H_0 and for any $x \in \Omega$, $\lambda_1^x \ge \lambda_2^x \ge \ldots$ are the eigenvalues of the covariance surface given by*

$$C^x(u, v) = \sqrt{w_x(u) w_x(v)}\, \operatorname{Cov}\big(\mathbb{I}\big(d\big(x, V'\big) \le u\big), \mathbb{I}\big(d\big(x, V'\big) \le v\big)\big)$$

with $V' \sim P$.

To study the asymptotic power of the proposed test, we consider a sequence of alternatives

$$\begin{aligned} H_{nm} = \big\{&(P_1, P_2) : X \sim P_1, Y \sim P_2, \\ &\text{with } D_{XY}^w = a_{nm}, a_{nm} \to 0, nm/(n+m)a_{nm} \to \infty, n, m \to \infty\big\}, \end{aligned} \tag{29}$$

with D_{XY}^w as in (26). The $\{H_{nm}\}$ form a sequence of contiguous alternatives shrinking toward H_0. The power of the test under this sequence is

$$\beta_{nm}^w = \mathbb{P}_{H_{nm}}\big(T_{nm}^w > q_\alpha\big), \tag{30}$$

where $q_\alpha = \inf\{t : \Gamma_L(t) \ge 1 - \alpha\}$ and $\Gamma_L(\cdot)$ is the cumulative distribution function of the asymptotic null distribution corresponding to the law of L in Theorem 6.1. Our next result shows that the proposed test is consistent against the contiguous alternatives $\{H_{nm}\}$ (29).

THEOREM 6.2. *Under Assumptions 1, 4, 5 and 6, for a sequence of alternatives H_{nm}, the power of the level α test (30) satisfies $\beta_{nm}^w \to 1$.*

Theorem 6.3 below provides theoretical guarantees for the permutation version of the test based on the empirical cut-offs for the randomization approximation of $\Gamma_L(\cdot)$ given by $\widehat{\Gamma}_{nm}(\cdot)$ (27). The consistency of the estimated critical value $\widehat{q}_\alpha$ (28) under H_0 is given by (32) and under alternatives $P_1 \ne P_2$ we consider a mixture distribution $\bar{P} = cP_1 + (1-c)P_2$ with $0 \le c \le 1$. Assume $\bar{X} = \{\bar{X}_1, \ldots, \bar{X}_n\}$ and $\bar{Y} = \{\bar{Y}_1, \ldots, \bar{Y}_m\}$ are i.i.d. samples from $\bar{P}$ and $T_{nm}^w(\bar{X}, \bar{Y})$ is the test statistic obtained using the samples $\bar{X}$ and $\bar{Y}$. We show in the proof of Theorem 6.3 in the Supplementary Material that under Assumptions 1, 4, 5 and 6, Theorem 6.1 can be utilized to obtain the asymptotic distribution of $T_{nm}^w(\bar{X}, \bar{Y})$ with cumulative distribution $\bar{\Gamma}_L(\cdot)$ and $\bar{q}_\alpha = \inf\{t \ge 0 : \bar{\Gamma}_L(t) \ge 1 - \alpha\}$. Suppose that $\bar{\Gamma}_L(\cdot)$ is continuous and strictly increasing at $\bar{q}_\alpha$ and $n, m \to \infty$ such that $\frac{n}{n+m} - c = O((n+m)^{-1/2})$ and $\frac{m}{n+m} - (1-c) = O((n+m)^{-1/2})$. Then under Assumptions 1, 4, 5 and 6, $|\widehat{q}_\alpha - \bar{q}_\alpha| = o_{\mathbb{P}}(1)$, that is, $\widehat{q}_\alpha$ converges to a deterministic limit also for the case where $P_1 \ne P_2$. This implies convergence of the power function $\tilde{\beta}_{nm}^w \to 1$ as $n, m \to \infty$, where

$$\tilde{\beta}_{nm}^w = \mathbb{P}_{H_{nm}}\big(T_{nm}^w > \widehat{q}_\alpha\big) \tag{31}$$

is the power function of the test under the sequence of the contiguous alternatives H_{nm} when using the permutation-derived critical value $\widehat{q}_\alpha$ instead of q_α.

THEOREM 6.3. *Under H_0 (20) and Assumptions 1, 4, 5 and 6, as $n, m \to \infty$ and $K \to \infty$ it holds that $|\widehat{\Gamma}_{m,n}(t) - \Gamma_L(t)| = o_{\mathbb{P}}(1)$ for every t which is a continuity point of $\Gamma_L(\cdot)$. Suppose that $\Gamma_L(\cdot)$ is continuous and strictly increasing at q_α. Then under H_0 (20) and Assumptions 1, 4, 5 and 6, as $n, m \to \infty$ and $K \to \infty$,*

$$|\widehat{q}_\alpha - q_\alpha| = o_{\mathbb{P}}(1). \tag{32}$$

Assume further that $\bar{\Gamma}_L(\cdot)$ is continuous and strictly increasing at $\bar{q}_\alpha$ and $\frac{n}{n+m} - c = O((n+m)^{-1/2})$ and $\frac{m}{n+m} - (1-c) = O((n+m)^{-1/2})$ as $n, m \to \infty$. Then under Assumptions 1, 4, 5 and 6 for the sequence of alternatives H_{nm}, the power (31) of the permutation test satisfies $\tilde{\beta}^w_{nm} \to 1$ as $n, m \to \infty$.

6.3. *Empirical experiments*. To illustrate the finite-sample performance of the proposed test, we performed simulation studies for various scenarios. Specifically, random objects included samples of random vectors with the Euclidean metric, samples of 2-dimensional distributions with the L^2 metric between corresponding cumulative distribution functions (cdfs), and samples of random networks from the preferential attachment model (Barabási and Albert (1999)) with the Frobenius metric between the adjacency matrices. In each scenario, we generated two samples of random objects of equal size $n = m = 100$ unless otherwise noted and performed 500 Monte Carlo runs to construct empirical power functions as the distance of the distributions of the first and second sample varies. The empirical power was assessed as the proportion of rejections of the test for the significance level 0.05 among the 500 Monte Carlo runs. We used the permutation version of the test and assessed p-values from $K = 1000$ permutations through the proportion of permutations yielding test statistics greater than the test statistics computed from the original sample. This proportion is $(K+1)^{-1}\{\sum_{j=1}^{K} \mathbb{I}(T_{\Pi_j} \geq T^w_{nm}) + 1\}$, with T^w_{nm} in (24) and T_{Π_j} for $j = 1, \ldots, K$ defined in the paragraph just before equation (27), where the case $j = 0$ corresponds to the original sample without permutation.

We compared the performance of the proposed test with the energy test (Székely and Rizzo (2004)) and the graph based test (Chen and Friedman (2017)). For the energy test, we obtained p-values based on 1000 permutations. For the graph based test, similarity graphs of all the observations pooling the two samples together were constructed as 5-MSTs, as suggested by Chen and Friedman (2017). Here, MST stands for minimum spanning tree, and a k-MST is the union of the 1st, …, kth MST(s), where a kth MST is a spanning tree connecting all observations while minimizing the sum of distances between connected observations subject to the constraint that all the edges are not included in the 1st, …, $(k-1)$th MST(s). In the scenarios with samples of multivariate data, we also included comparisons of the proposed test with the two-sample Hotelling's T^2 test.

In the following figures illustrating power comparisons, "energy" stands for the energy test (Székely and Rizzo (2004)), "graph" for the graph based test (Chen and Friedman (2017)), "Hotelling" for the two-sample Hotelling's T^2 test and "DP" for the proposed distance profile-based test (23). For samples of multivariate data endowed with the Euclidean metric, we generated the data in four scenarios. In the first two scenarios, we generated two samples of p-dimensional random vectors $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^m$ from a Gaussian distribution $N(\mu, \Sigma)$ for dimensions $p = 30$, 90 and 180, respectively. In the first scenario, the population distributions of random vectors in the two samples differ only in the mean μ while the population covariance matrix $\Sigma = U \Lambda U^\top$ is the same for both samples, where Λ is a diagonal matrix with kth diagonal entry $\cos(k\pi/p) + 1.5$ for $k = 1, \ldots, p$ and U is an orthogonal matrix with

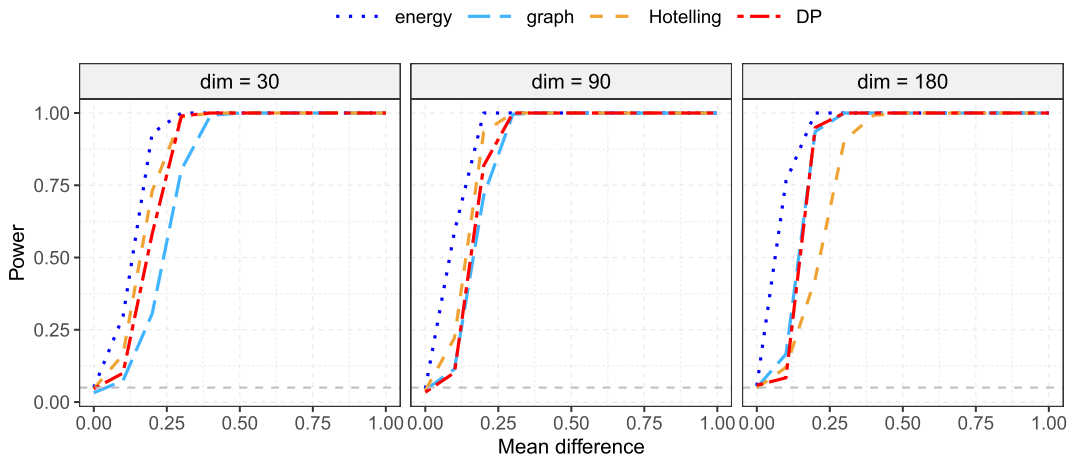

FIG. 3. *Power comparison for increasing values of mean difference* Δ_1 *for two samples of* p*-dimensional random vectors sampled from* $N(\mu, \Sigma)$*. Here,* $\mu = \mathbf{0}_p = (0, 0, \ldots, 0)^\top$ *for the first samples and* $\mu = \Delta_1 \mathbf{1}_p = \Delta_1(1, 1, \ldots, 1)^\top$ *for the second samples;* $\Sigma = U \Lambda U^\top$ *for both samples, where* Λ *is a diagonal matrix with* k*th diagonal entry* $\cos(k\pi/p) + 1.5$ *for* $k = 1, \ldots, p$ *and* U *is an orthogonal matrix with first column* $p^{-1/2}(1, 1, \ldots, 1)^\top$*. The dashed grey line denotes the significance level* 0.05.

first column $p^{-1/2}(1, 1, \ldots, 1)^\top$. Specifically, $\mu = \mathbf{0}_p = (0, 0, \ldots, 0)^\top$ for the first samples $\{X_i\}_{i=1}^n$, and $\mu = \Delta_1 \mathbf{1}_p = \Delta_1(1, 1, \ldots, 1)^\top$ for the second samples $\{Y_i\}_{i=1}^m$, where Δ_1 ranges from 0 to 1. The results are shown in Figure 3. In addition, we considered another location shift scenario for lower-dimensional data; see Section S.5 in the Supplementary Material for details.

In the second scenario, the population distributions of the two samples differ only in scale, while sharing the same mean $\mu = \mathbf{0}_p$. For the first samples $\{X_i\}_{i=1}^n$, $\Sigma = 0.8 I_p$ and for the second samples $\{Y_i\}_{i=1}^m$, $\Sigma = (0.8 - \Delta_2) I_p$ with Δ_2 ranging from 0 to 0.4. The results are shown in Figure 4. In the first scenario with location shifts, the proposed test outperforms the graph based test when the dimension is relatively low ($p = 30$); the graph based test catches up when the dimension is high. Meanwhile, the energy test is always the winner in this scenario. The performance of Hotelling's T^2 test is the second best when the dimension p is

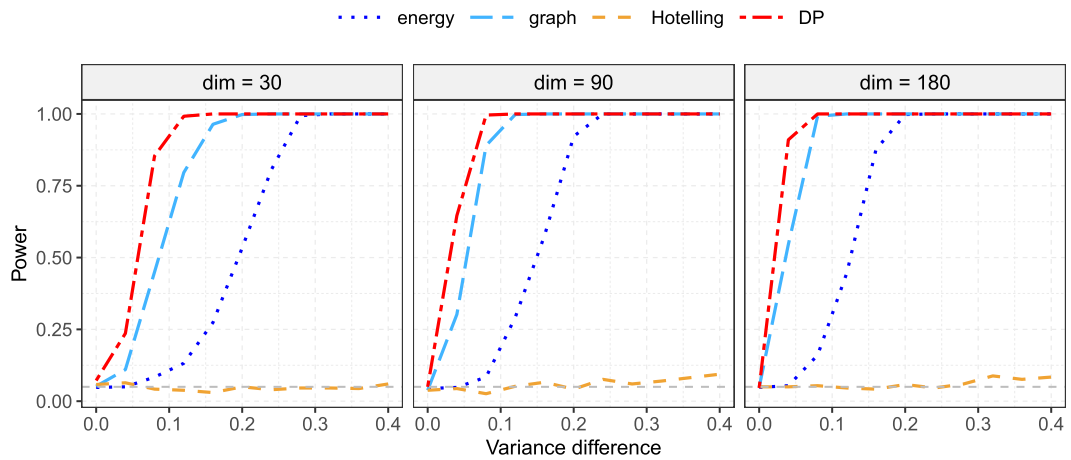

FIG. 4. *Power comparison for increasing values of variance difference* Δ_2 *for two samples of* p*-dimensional random vectors sampled from* $N(\mu, \Sigma)$*. Here,* $\mu = \mathbf{0}_p$ *for both samples;* $\Sigma = 0.8 I_p$ *for the first samples and* $\Sigma = (0.8 - \Delta_2) I_p$ *for the second samples. The dashed grey line denotes the significance level* 0.05.

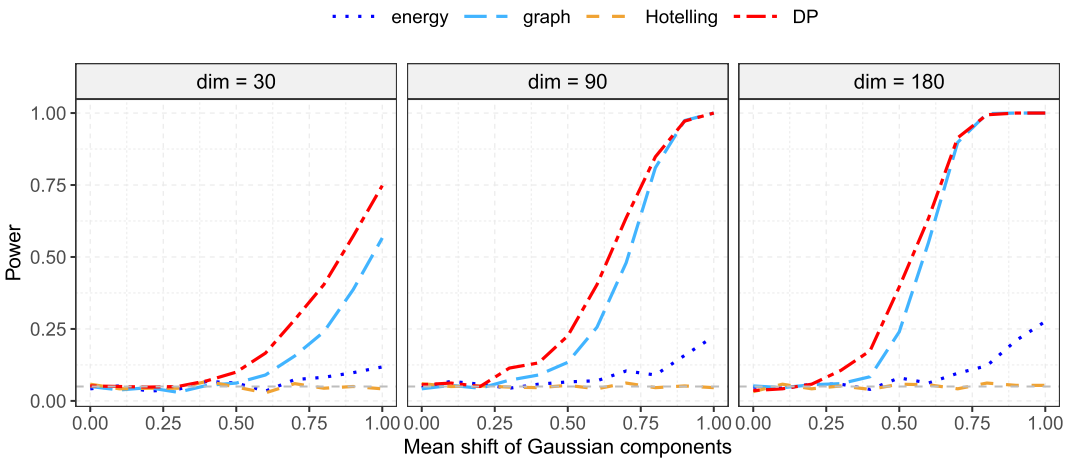

FIG. 5. *Power comparison for increasing values of mean shift Δ_3 of Gaussian components for two samples of p-dimensional random vectors. Here, the first samples are sampled from $N(\mathbf{0}_p, I_p)$; the second samples consist of $AZ_1 + (1-A)Z_2$, where $A \sim \text{Bernoulli}(0.5)$, $Z_1 \sim N(-\mu, I_p)$, $Z_2 \sim N(\mu, I_p)$, $\mu = (\Delta_3 \mathbf{1}_{0.1p}^\top, \mathbf{0}_{0.9p}^\top)^\top$, and A, Z_1 and Z_2 are independent. The dashed grey line denotes the significance level* 0.05.

less than the sample size $n = m = 100$ but drops to the bottom when $p = 180$ and $\Delta_1 > 0.1$. In the second scenario with scale changes, we find that the proposed test always outperforms all the other tests.

In the third scenario, the first samples of random vectors $\{X_i\}_{i=1}^n$ are generated from the Gaussian distribution $N(\mathbf{0}_p, I_p)$, and the second samples of random vectors $\{Y_i\}_{i=1}^m$ are generated from a mixture of two Gaussian distributions with the overall population mean equaling that of the first samples. Specifically, the second samples consist of independent copies of $AZ_1 + (1-A)Z_2$, where $A \sim \text{Bernoulli}(0.5)$, $Z_1 \sim N(-\mu, I_p)$, $Z_2 \sim N(\mu, I_p)$, $\mu = (\Delta_3 \mathbf{1}_{0.1p}^\top, \mathbf{0}_{0.9p}^\top)^\top$, and A, Z_1 and Z_2 are independent. Here, Δ_3 ranges from 0 to 1. The results are shown in Figure 5; again, the proposed test outperforms all the other tests.

The fourth scenario compares Gaussian distributions with heavy-tailed distributions, where the first samples $\{X_i\}_{i=1}^n$ are generated from $N(\mathbf{0}_p, I_p)$, and the second samples $\{Y_i\}_{i=1}^m$ consist of random vectors with components that are independent and identically distributed following a t distribution with degrees of freedom ranging from 2 to 22. The results for $p \in \{5, 15, 60\}$ are in Figure 6; the proposed test outperforms all the other tests.

Next, we considered bivariate probability distributions as random objects, where we use the L^2 distance between corresponding cdfs as the metric between two probability distributions. Each observation X_i or Y_i is a random 2-dimensional Gaussian distribution $N(Z, 0.25I_2)$, where Z is a 2-dimensional random vector, with two scenarios: In the first scenario, $Z \sim N(\mathbf{0}_2, 0.25I_2)$ for the first samples and $Z \sim N((\delta_1, 0)^\top, 0.25I_2)$ for the second samples. In the second scenario, $Z \sim N(\mathbf{0}_2, 0.4^2 I_2)$ for the first samples and $Z \sim N(\mathbf{0}_2, \text{diag}((0.4 + \delta_2)^2 I_2))$ for the second samples. The results are presented in Figures 7 and 8, respectively. The first scenario showcases location shifts of Z; the proposed test outperforms the graph based test but is outperformed by the energy test. The second scenario showcases scale changes of Z, where the proposed test outperforms all the other tests.

We also studied the power of the proposed test for random networks endowed with the Frobenius metric between adjacency matrices as random objects. Each datum X_i or Y_i is a random network with 200 nodes generated from the preferential attachment model (Barabási and Albert (1999)) with the attachment function proportional to k^γ, where $\gamma = 0$ for the first samples $\{X_i\}_{i=1}^n$, and γ increases from 0 to 0.5 for the second samples $\{Y_i\}_{i=1}^m$. As shown in Figure 9, the proposed test outperforms both the energy test and the graph based test. In

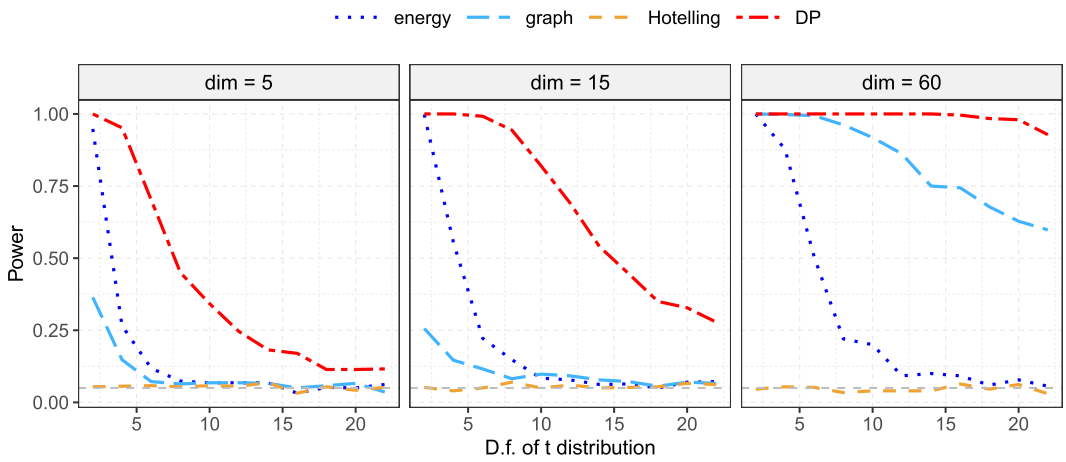

FIG. 6. *Power comparison for increasing values of ν for two samples of p-dimensional random vectors. Here, the first samples are sampled from $N(\mathbf{0}_p, I_p)$; the second samples consist of random vectors with independent components, where each component follows a t-distribution with ν degrees of freedom (d.f.). The dashed grey line denotes the significance level* 0.05.

addition to the case with $n = m = 100$, we performed simulations with larger samples of sizes $n = m = 200$. The corresponding results are shown in Figure S.2 in Section S.5 in the Supplementary Material and they more or less match those for $n = m = 100$.

7. Extensions and data illustrations.

7.1. *Profile metric and object data visualization.* Distance profiles induce a new similarity measure in Ω that we refer to as *profile metric* d_P. It complements the original metric d and depends on d, the underlying probability measure P and also on the distributional metric in the space of distance profiles, for which we select the Wasserstein metric. The profile metric quantifies the distance of the profile densities of elements of Ω,

$$d_P(\omega_1, \omega_2) = d_W(F_{\omega_1}, F_{\omega_2}), \tag{33}$$

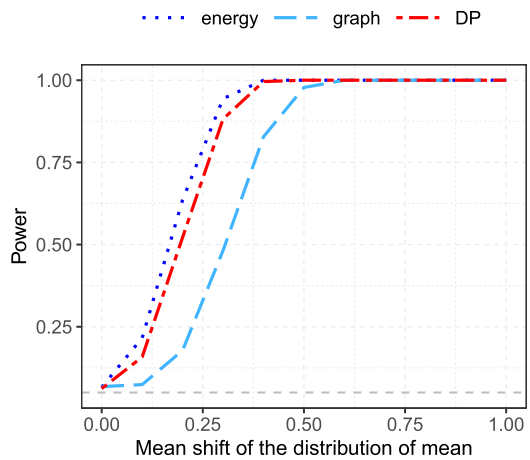

FIG. 7. *Power comparison for increasing values of mean shift δ_1 of the distribution of the mean for two samples of random bivariate Gaussian distributions $N(Z, 0.25I_2)$, where $Z \sim N(\mathbf{0}_2, 0.25I_2)$ for the first samples and $Z \sim N((\delta_1, 0)^\top, 0.25I_2)$ for the second samples. The dashed grey line denotes the significance level* 0.05.

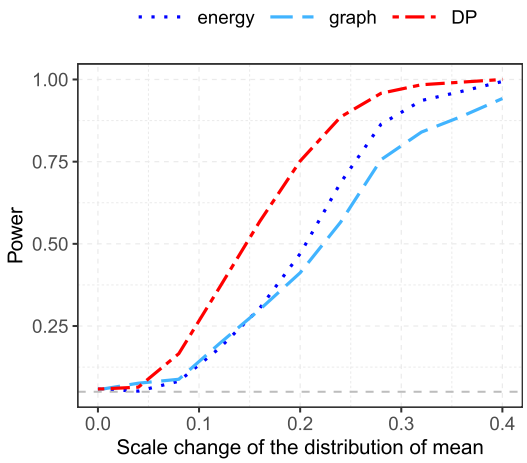

FIG. 8. *Power comparison for increasing values of scale change δ_2 of the distribution of the mean for two samples of random bivariate Gaussian distributions $N(Z, 0.25I_2)$, where $Z \sim N(\mathbf{0}_2, 0.4^2 I_2)$ for the first samples and $Z \sim N(\mathbf{0}_2, \mathrm{diag}((0.4+\delta_2)^2 I_2)$ for the second samples. The dashed grey line denotes the significance level* 0.05.

where d_W is the Wasserstein metric (1) and F_{ω_1}, F_{ω_2} are the distance profiles of ω_1, ω_2, as defined in (8). It is not a genuine metric on Ω but rather a measure of dissimilarity of the distance profiles of elements of Ω.

The profile metric d_P generally may differ substantially from the original metric d. For example, two outlying elements of Ω may be far away from each other in terms of the original metric d but if they have similar centrality and distance profiles they will have small profile dissimilarity d_P, which could be 0 if their distance profiles coincide. It turns out that the profile metric is very useful for data analysis, as we will demonstrate in the following. Its implementation depends on distance profiles, which must be estimated from the available data, and thus the profile metric itself is only available in the form of an estimate.

To visualize random objects, low-dimensional projections of similarities as afforded by MDS are a prime tool and any MDS version (Mardia (1978)) can be based on either the original distance d, in the following referred to as *object MDS* or alternatively on the profile

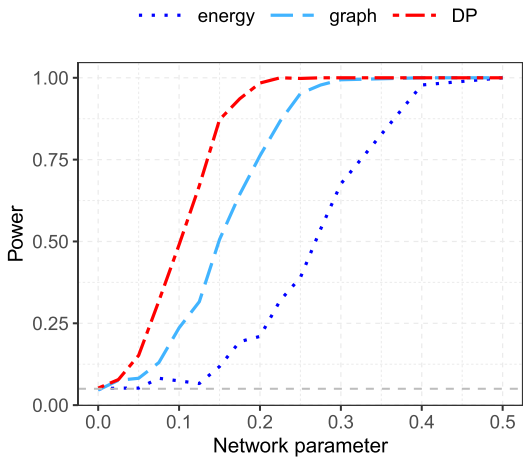

FIG. 9. *Power comparison for increasing values of γ for two samples of random networks with* 200 *nodes from the preferential attachment model. The attachment function is proportional to k^γ with $\gamma = 0$ for the first samples, and γ increasing from* 0 *to* 0.5 *for the second samples. The dashed grey line denotes the significance level* 0.05.

metric, in the following referred to as *profile MDS*. In profile MDS, we use the estimated distance profiles $\widehat{F}_{X_i}$ and the Wasserstein metric d_W (1), while we use the distance d in Ω for object MDS. In the following, MDS is implemented with `cmdscale()` in the R built-in package `stats` (R Core Team (2020)).

To enhance the graphical illustration of the proposed transport ranks in (12), for implementations, data applications and simulations we found that partitioning the observed random objects into a not too large number of k groups according to their transport ranks is advantageous for visualization and communicating results. Specifically, the range of the transport ranks of observations within a sample $\{X_i\}_{i=1}^n$ is partitioned into k bins, $S_k = [0, q_{1/k}]$, $S_{k-1} = (q_{1/k}, q_{2/k}], \ldots, S_1 = (q_{(k-1)/k}, 1]$, where q_α is the α-quantile of $\{\widehat{R}_{X_i}\}_{i=1}^n$ for $\alpha \in (0, 1)$; then the jth group consists of observations with transport ranks falling in S_j for $j = 1, \ldots, k$. Arranging the bins in descending order of transport ranks, these groups are ordered from the innermost to the outermost, providing a center-outward description of the data; we found that the choice $k = 10$ worked well, as illustrated in Figures 10–13 below. The function `CreateDensity()` in the R package `frechet` (Chen et al. (2020)) was used to obtain Wasserstein barycenters of distance profiles for each group.

The code for obtaining distance profiles, transport ranks, object MDS plots and profile MDS plots is available on GitHub (Chen, Dubey and Müller (2024)).

7.2. *Illustrations with simulated data.* We start with a simple special case of a Euclidean vector space, where we sampled $n = 500$ observations $\{X_i\}_{i=1}^n$ independently from a p-dimensional Gaussian distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for $p = 2$ and $p = 50$, with $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \operatorname{diag}(p, p-1, \ldots, 1)$. The distance profiles $\widehat{F}_{X_i}$ (15) and transport ranks $\widehat{R}_{X_i}$ (16) were computed for each observation, adopting the Euclidean metric in $\mathbb{R}^p$. Irrespective of the type of random objects X_i, the distance profiles $\widehat{F}_{X_i}$ are situated in the space $\mathcal{W}$ of one-dimensional distributions with finite second moments with the Wasserstein metric (1).

For $p = 2$, the transport ranks (16) based on distance profiles capture the center-outward ordering of the 2-dimensional Gaussian data and the Wasserstein barycenters of the distance profiles within each group shift to the right from group 1 to group 10, where the grouping is as described in Section 7.1, reflecting increased distances from the bulk of data (Figure 10). Figure 11 demonstrates profile MDS for a simulated sample of $n = 500$ observations from a 50-dimensional Gaussian distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \operatorname{diag}(50, 49, \ldots, 1)$ and shows that profile MDS provides a simple representation by sorting these high-dimensional Euclidean data along dimension 1.

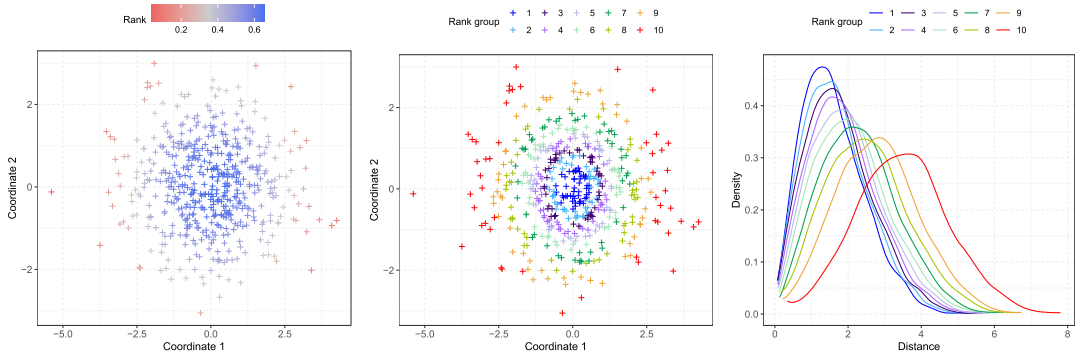

FIG. 10. *Scatterplots of a sample of $n = 500$ observations generated from a 2-dimensional Gaussian distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \operatorname{diag}(2, 1)$, where the points are colored according to their transport ranks (16) (left) and grouped into 10 groups according to the quantiles of transport ranks (middle); Wasserstein barycenters of the distance profiles within each group represented by density functions (right).*

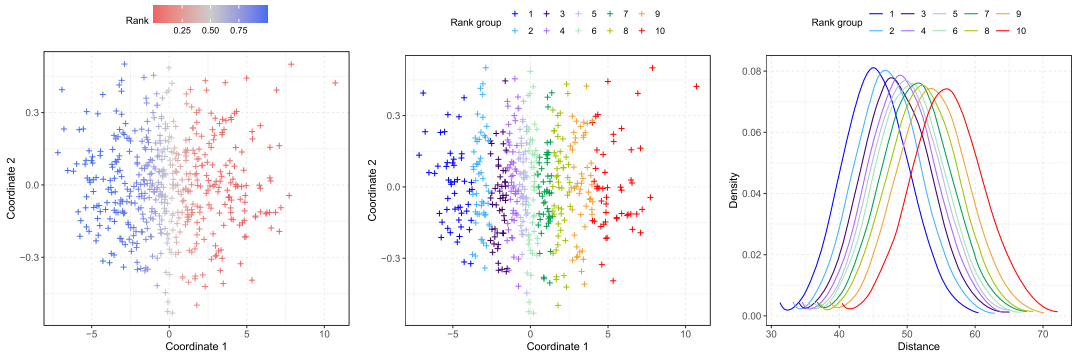

FIG. 11. *Two-dimensional (profile) MDS with respect to the Wasserstein metric d_W in (1) of the distance profiles $\widehat{F}_{X_i}$ (15) with $\omega = X_i$ of a sample of $n = 500$ observations generated from a 50-dimensional Gaussian distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \operatorname{diag}(50, 49, \ldots, 1)$, where the points are colored according to their transport ranks (16) (left) and grouped into 10 groups according to the quantiles of transport ranks (middle); Wasserstein barycenters of the distance profiles within each group (right).*

Additional simulation results can be found in the Supplementary Material for random objects corresponding to 2-dimensional random vectors generated from multimodal distributions in Section S.6 and for distributional data in Section S.7.

7.3. *Illustration with human mortality data.* Understanding human longevity has been of long-standing interest and age-at-death distributions are relevant random objects for this endeavor. We consider age-at-death distributions for different countries, obtained from the Human Mortality Database (http://www.mortality.org) for the year 2000 for $n = 34$ countries, separately for males and females. The age-at-death distributions are shown in the form of density functions in Figure S.5 in Section S.8 in the Supplementary Material. To analyze the data geometry of this sample of random distributions $\{X_i\}_{i=1}^{34}$, we assume that they are situated in a space (Ω, d) of distributions equipped with the Wasserstein metric $d = d_W$ (1) and then obtained distance profiles $\widehat{F}_{X_i}$ (15) for $\omega = X_i$ for each country.

In Figure 12, we compare profile MDS (based on the distance of profiles, where the Wasserstein metric is applied for the distributional space where the profiles are situated) in the left panels and object MDS (based on the original metric d in the object space of distributions); mortality for females is shown in the top panels and for males in the bottom panels. We find that profile MDS leads to a clearly interpretable one-dimensional manifold representation for both females and males, where extremes appear at each end, at the red colored end corresponding to age-at-death distributions indicating reduced and at the green colored end enhanced longevity. The groups of countries that form the extreme ends are Japan at the enhanced and Eastern European countries, such as Russia, Ukraine, Belarus, Latvia and Estonia at the reduced longevity end. Luxembourg and Belgium belong to the most central group for both females and males. Spain is among the more outlying countries for females only, with a longevity increase for females but not for males. One can observe many other interesting features in terms of the similarity and contrast between females' and males' longevity for specific countries, for example, for Denmark and Netherlands. We find that the one-dimensional ordering provided by profile MDS facilitates the interpretation and communication of the main data features, while object MDS is less informative.

Another finding of interest that emerges from profile MDS is that the age-at-death distributions for males for the outlying countries are more outlying than the corresponding age-at-death distributions for females. In particular, the empirical Fréchet variances of the distance profiles, $n^{-1}\sum_{i=1}^{n} d_W^2(\widehat{F}_{X_i}, \widehat{F}_{\oplus})$, of age-at-death distributions for females and males of different countries are 2.08 and 8.22, respectively, where $\widehat{F}_{\oplus} = \operatorname{argmin}_{\omega \in \mathcal{W}} \sum_{i=1}^{n} d_W^2(\widehat{F}_{X_i}, \omega)$

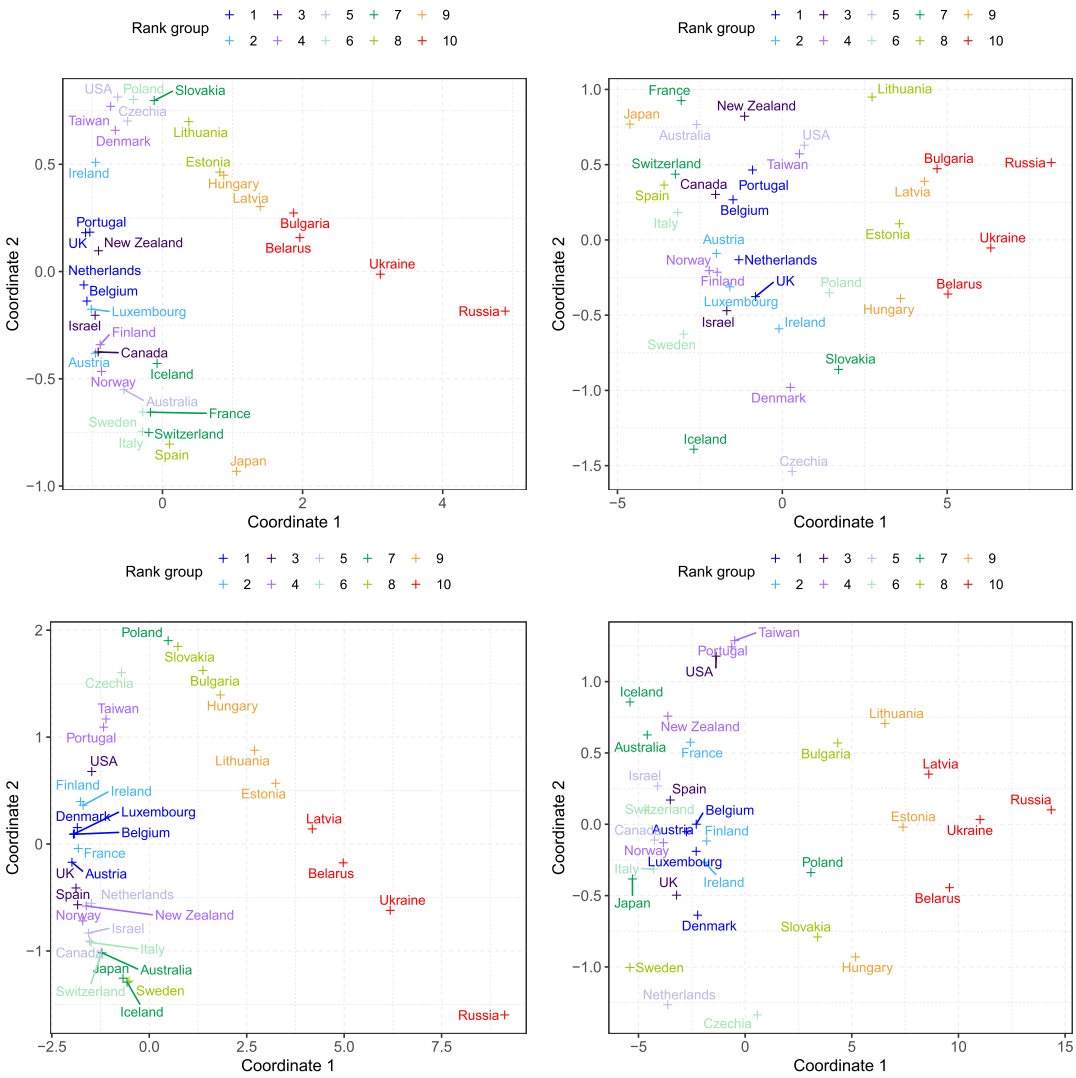

FIG. 12. *Two-dimensional profile MDS (left) and object MDS (right) of the age-at-death distributions of females (top) and males (bottom) in* 2000 *for* 34 *countries, where the objects are grouped into* 10 *groups and colored according to the quantiles of their transport ranks.*

is the empirical Fréchet mean of the distance profiles, indicating that male age-at-death is especially sensitive to unfavorable country-specific factors such as the lingering effects of societal upheaval in Eastern Europe.

7.4. *Illustration with U.S. electricity generation data.* Compositional data comprise another type of data that do not lie in a vector space. Such data are commonly encountered and consist of vectors of nonnegative elements that sum up to 1. Examples include geochemical compositions and microbiome data. Various approaches to handle the nonlinearity that is inherent in such data have been developed (Aitchison (1986), Filzmoser, Hron and Templ (2019), Scealy and Welsh (2014)). We consider here the U.S. electricity generation data, which are publicly available on the website of the U.S. Energy Information Administration (http://www.eia.gov/electricity). The data consist of net generation of electricity from different sources for each state. We considered the data for the year 2000. In preprocessing, we excluded the "pumped storage" category due to errors in these data and then merged the

other energy sources into three categories: Natural Gas, consisting of "natural gas" alone; Other Fossil, consisting of "coal," "petroleum" and "other gases"; Renewables and Nuclear, combining the remaining sources "hydroelectric conventional," "solar thermal and photovoltaic," "geothermal," "wind," "wood and wood derived fuels," "other biomass," "nuclear" and "other." Hence, we have a sample of $n = 50$ observations $\{X_i\}_{i=1}^n$, each of which takes values in a 2-simplex $\Delta^2 = \{\boldsymbol{x} \in [0,1]^3 : \boldsymbol{x}^\top \mathbf{1}_3 = 1\}$, where $\mathbf{1}_3 = (1,1,1)^\top$. Since the componentwise square root $\sqrt{\boldsymbol{x}} = (\sqrt{x_1}, \sqrt{x_2}, \sqrt{x_3})^\top$ of an element $\boldsymbol{x} \in \Delta^2$ lies in the unit sphere $\mathcal{S}^2$, we adopted the geodesic metric on this sphere

$$d_S(\boldsymbol{x}, \boldsymbol{y}) = \arccos\big(\sqrt{\boldsymbol{x}}^\top \sqrt{\boldsymbol{y}}\big) \quad \text{for } \boldsymbol{x}, \boldsymbol{y} \in \Delta^2. \tag{34}$$

We then compared the proposed transport ranks (16) for each state with the angular Tukey depths (ATDs, Liu and Singh (1992)) of $\{\sqrt{X_i}\}_{i=1}^n$. At first glance, the proposed transport ranks and ATDs yield similar center-outward ordering of the 50 states for these data (Figure 13). Maryland emerges as the transport median and is also at the median in terms of ATDs. On closer inspection, one finds some interesting discrepancies between transport ranks and the ATDs, especially for the states that are either close to or far away from the center Maryland in terms of their outlyingness. The states near Maryland, as shown in orange and light violet in the bottom panels of Figure 13, all have high transport ranks, while their ATDs vary widely. In particular, Montana, with an electricity generation pattern very similar to that of Maryland, has the lowest ATD level while it has a high transport rank. A subset of states that are colored in turquoise and light violet in the bottom panels of Figure 13 have the lowest ATDs among all states but have a much wider range of transport ranks. For example, Hawaii and Delaware for which energy sources are similar to those of Maryland have high transport ranks and low ATD levels. The overall conclusion is that transport ranks are better suited than ATDs for studying the geometry of this data set and for quantifying outlyingness.

Networks as random objects are illustrated in another data application for New York City taxi trips; details can be found in Section S.9 of the Supplementary Material.

7.5. *Illustrations of the two-sample test.*

7.5.1. *Human mortality data.* We illustrate the proposed two-sample test with the age-at-death distributions from the Human Mortality Database as described in Section 7.3. The countries we considered are Belarus, Bulgaria, Czechia, Estonia, Hungary, Latvia, Lithuania, Poland, Russia, Slovakia and Ukraine, which are all Eastern European countries at the lowest longevity levels. One question of interest is whether the age-at-death distributions of these Eastern European countries changed after the dissolution of the Soviet Union.

To this end, we compared the age-at-death distributions in 1990 and the distributions in 1993 for these countries separately for females and males, utilizing the proposed test, as well as the energy test (Székely and Rizzo (2004)) and the graph based test (Chen and Friedman (2017)). The densities of these distributions are shown in Figure 14 and the test results are summarized in Table 1, where the tests are implemented and referred to in the same way as in the simulations in Section 6.3 and p-values less than 0.05 are highlighted in bold.

In Figure 14, it can be seen that the age-at-death densities of males in 1993 vary more across the Eastern European countries than in 1990 while the variation of those for females is more similar. While the proposed test does not find a significant difference between age-at-death distributions for females in 1990 and 1993, the p-value of the proposed test for males is below 0.05, which provides evidence that a systematic change occurred in the age-at-death distributions for males in these Eastern European countries between 1990 and 1993. In contrast, the energy test and the graph based test do not find significant differences at the $\alpha = 0.05$ significance level for either females or males.

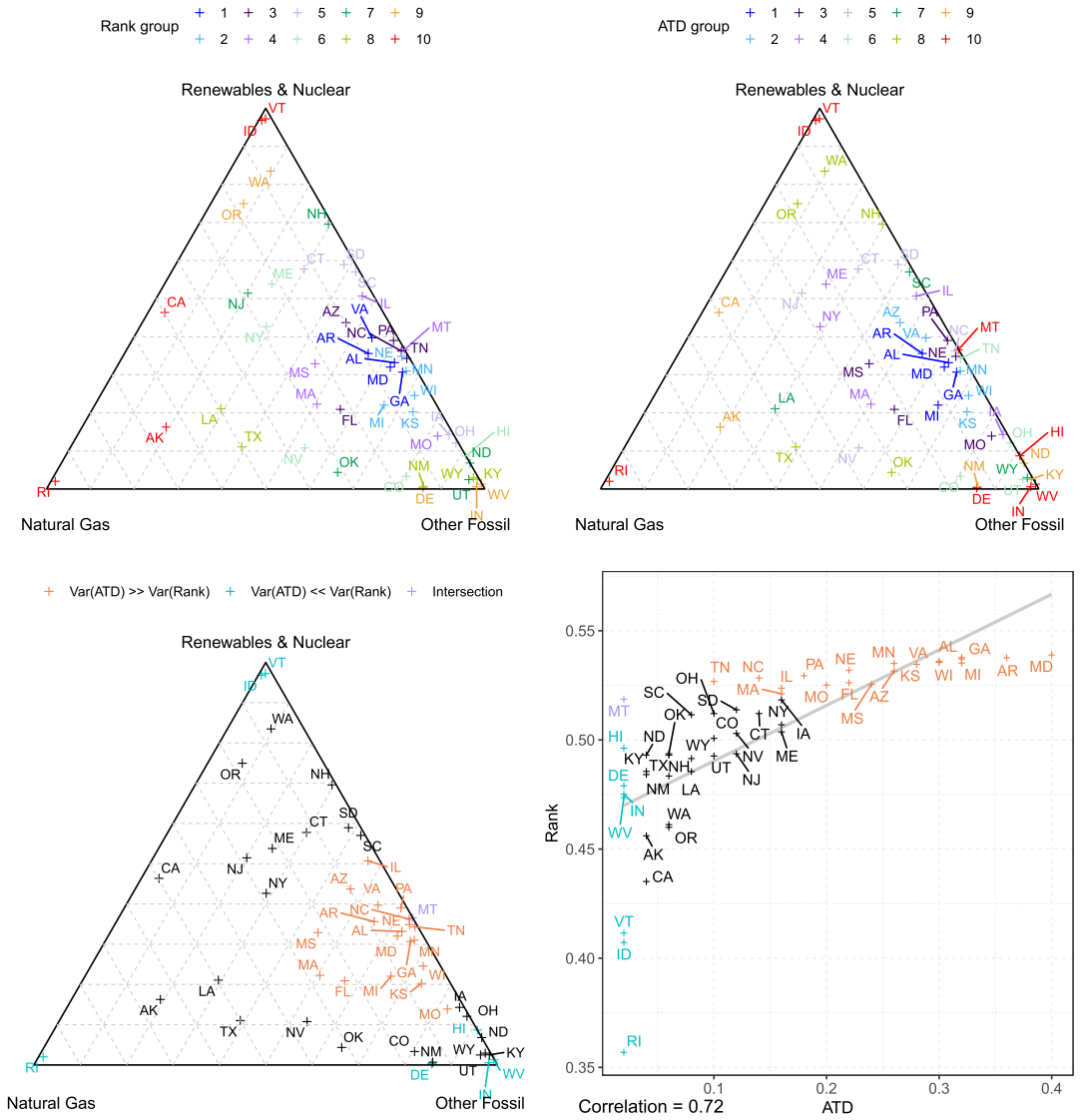

FIG. 13. *Ternary plot of compositions of electricity generation in the year* 2000 *for the* 50 *states in the U.S., where the points are colored as per their grouping into* 10 *groups according to quantiles of transport ranks* (*top left*); *the corresponding grouping based on angular Tukey depths* (*ATDs*, *top right*); *highlighted subsets that show differences between transport ranks and ATDs* (*bottom left*); *and a scatterplot of transport ranks* (16) *versus ATDs* (*bottom right*), *where the straight line shows the least squares fit* (*to provide perspective*). *In the bottom two panels, a subset of states with similarly small ATDs but varying transport ranks is highlighted in turquoise, and another subset with similarly high transport ranks but varying ATDs is highlighted in orange, where the intersection of these two subsets is colored in light violet.*

7.5.2. *Functional connectivity networks based on fMRI data.* Functional connectivity in neuroimaging refers to temporal association of a neurophysiological measure obtained from different regions in the brain (Friston et al. (1993)). Functional magnetic resonance imaging (fMRI) techniques record time courses of blood oxygenation level dependent (BOLD) signals, which are a proxy for neural activity in the brain (Lindquist (2008)). Specifically, resting state fMRI (rs-fMRI) records signals when subjects are resting and not performing an explicit task. Functional connectivity networks can be constructed across various brain regions of interest (ROIs) by applying a threshold to certain measures of temporal association for each pair of ROIs that in an initial step are represented as symmetric correlation matrices.

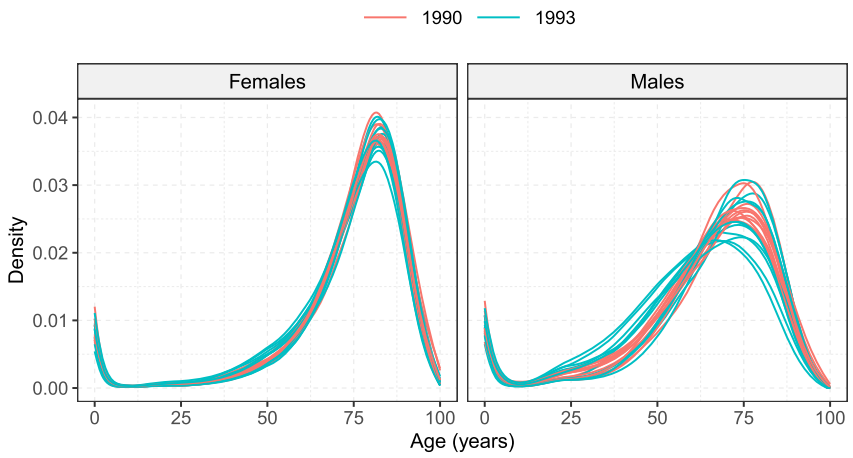

FIG. 14. *Age-at-death distributions for females and males in the eleven Eastern European countries in* 1990 (*red*) *and* 1993 (*blue*), *all shown as density functions*.

The rs-fMRI data in our analysis were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu), including 400 clinically normal (CN) subjects and 85 mild Alzheimer's disease dementia (AD) subjects. For each subject, we took only their first scan. Preprocessing of the BOLD signals was implemented following the standard procedures of head motion correction, slice-timing correction, coregistration, normalization and spatial smoothing. Average signals of voxels within spheres of diameter 8 mm centered at the seed voxels of each ROI were extracted, with linear detrending and band-pass filtering to account for signal drift and global cerebral spinal fluid and white matter signals, including only frequencies between 0.01 and 0.1 Hz. These steps were performed in MATLAB using the Statistical Parametric Mapping (SPM12, http://www.fil.ion.ucl.ac.uk/spm) and Resting-State fMRI Data Analysis Toolkit V1.8 (REST1.8, http://restfmri.net/forum/?q=rest).

We considered the 264 ROIs of a brain-wide graph identified by Power et al. (2011) and use temporal Pearson correlations (PCs) (Biswal et al. (1995)) as the measure of temporal correlation between pairs of ROIs, a common approach in fMRI studies. Functional connectivity networks were then obtained as adjacency matrices by imposing an absolute threshold 0.25 on the 264×264 matrices of temporal PCs, where entries less than 0.25 are replaced with zeros and diagonal entries are set to zero. As distance between two functional connectivity networks, we chose the Frobenius metric between the adjacency matrices.

We then employed the proposed test, the energy test and the graph based test to compare the functional connectivity networks of CN subjects and AD subjects. Prior to performing the tests, we subsampled the CN and AD subjects such that the age distributions of these two groups of subjects are similar. The results are presented in Table 2. The proposed test and

TABLE 1

p-values for testing whether the age-at-death distributions in 1990 *and the distributions in* 1993 *have the same distribution, for females and males, respectively*

Test	Females	Males
energy	0.326	0.089
graph	0.107	0.055
DP	0.159	**0.044**

TABLE 2
p-values for testing whether the functional connectivity networks of CN subjects and AD subjects have the same distribution

Test	p-value
energy	**0.003**
graph	0.930
DP	**0.034**

the energy test have p-values below 0.05, providing evidence for a significant differences between the distributions of functional connectivity networks of CN subjects and AD subjects, while the p-value of the graph based test is close to 1.

In a second analysis, we compared the functional connectivity networks of CN subjects for (first) scans taken at various age groups, with their distribution across age groups summarized in Table 3. Empirical power was obtained as the proportion of rejections at significance level $\alpha = 0.05$ based on 100 Monte Carlo runs, for each of a sequence of tests. For all tests, the first sample consisted of functional connectivity networks of 80 subjects randomly sampled from the 159 CN subjects with scans taken in the age interval $[55, 70)$. The second samples were drawn from the remaining 320 CN subjects and consisted of functional connectivity networks of subjects with scans taken in defined age intervals. For the first test, this age interval was $[55, 70)$, for the second test it was $[60, 75), \ldots,$ for the second-to-last test it was $[75, 90)$ and for the last test it was $[80, 96)$. Since it is known that these networks change with age, this sequence of tests provides an empirical power function for detecting the age-related change. The empirical power results are presented in Figure 15, indicating that the proposed test outperforms both the energy test and the graph based test.

8. Discussion and outlook.

8.1. *Metric selection.* To deploy the tools of metric statistics for a given space of data objects, the choice of a metric is essential. For some data types such as Euclidean data, the metric is usually preordained to be the geodesic, that is, the usual Euclidean metric, but even in this simple special case there are still various choices; one could consider weighted metrics that deemphasize or emphasize specific vector components. Similarly, for data on Riemannian manifolds such as spheres, the geodesic metric is an inherent feature of the geometry and, therefore, is the canonical choice. This applies also to compositional data if they are represented on the positive orthant of a unit sphere (Scealy and Welsh (2011)), as described in

TABLE 3
Age distribution at first scans for the CN subjects

Age interval	# CN subjects
$[55, 60)$	14
$[60, 65)$	20
$[65, 70)$	125
$[70, 75)$	84
$[75, 80)$	76
$[80, 85)$	49
$[85, 90)$	21
$[90, 95)$	10
$[95, 96)$	1

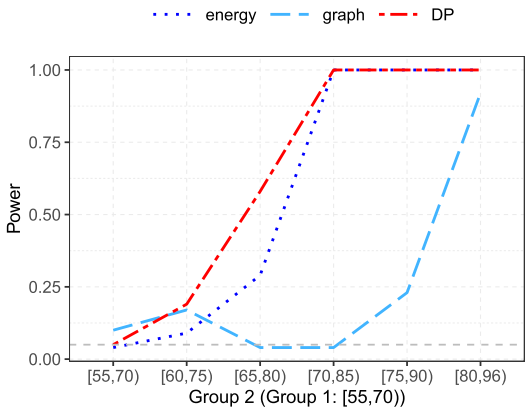

FIG. 15. *Power comparison for increasing age difference for two samples of functional connectivity networks of CN subjects, where the first samples consist of CN subjects with scans taken in the age interval* $[55, 70)$ *years and the second samples consist of CN subjects with scans taken in the age intervals* $[55, 70), [60, 75), \ldots, [75, 90), [80, 96)$, *respectively. For each Monte Carlo run, the first sample consists of* 80 *subjects which are randomly sampled from the* 159 *CN subjects with ages in* $[55, 70)$, *and from among the remaining* 320 *CN subjects, the second sample consists of the subjects with scans taken in the corresponding age interval. The dashed grey line denotes the significance level* 0.05.

Section 7.4; an alternative selection for compositional data is the Aitchison metric (Aitchison (1986)). Both choices have specific advantages and disadvantages (Scealy and Welsh (2014)), notably the Aitchison metric requires the compositional components to be positive (otherwise requiring a numerical fudge), which is not satisfied for the U.S. energy generation data that we study in Section 7.4.

The metric selection problem is more complex for other data objects such as distributions, where a large number of metrics have been proposed and popular choices in the context of random objects include the 2-Wasserstein metric (1). The Wasserstein metric has been shown to work well for one-dimensional distributions in distributional data analysis (Panaretos and Zemel (2020), Petersen, Zhang and Kokoszka (2022)) and various applied scenarios (Bolstad et al. (2003)), but poses thorny theoretical and computational problems for multivariate distributions. This incentivizes the study of alternative metrics such as the sliced Wasserstein metric (Chen and Müller (2023), Kolouri et al. (2019), Kolouri, Zou and Rohde (2016)) and the Fisher–Rao metric (Dai (2022)). For the space of symmetric positive matrices that play an important role for applications such as fMRI and DTI brain imaging, one can choose among a large class of metrics, ranging from the Frobenius metric to power metrics (Pigoli et al. (2014)), the recently proposed Cholesky metric (Lin (2019)) and metrics that reflect the geometry of eigenvectors (Jung, Schwartzman and Groisser (2015)).

While it is clearly important, the metric selection problem in a statistical framework has been largely neglected. If one has a class of metrics that is indexed by a parameter, a data-based selection criterion to find the best metric within the class may lead to consistent data-based metric selection for a specific target criterion; an example is metric selection for the family of power metrics for symmetric positive matrices (Petersen and Müller (2016b)). Absent a statistical framework for metric selection, a basic criterion is that the metric should be of strong negative type (9), which then implies that distance profiles characterize the underlying distribution. This property is satisfied for the metrics that we have discussed in examples, simulations and data analysis. We refer to Section 3 of Lyons (2013) for a detailed discussion of examples and counterexamples of metrics that are of (strong) negative type. Specifically, the space of one-dimensional distributions with finite second moments endowed with the 2-Wasserstein metric in equation (1), the space of multivariate distributions with

L^2 metric between corresponding cdfs, the space of network adjacency matrices with the Frobenius metric, and spheres with the geodesic metric are of strong negative type, while Grassmannian manifolds and cylinders with their geodesic metrics are not (Feragen, Lauze and Hauberg (2015), Venet (2019)).

Other criteria for metric selection include feasibility and ease of implementation (e.g., the Fisher–Rao metric can be easily deployed irrespective of the dimensionality of distributions) and matching of metric proximity with perceived or known similarity. A good metric should also facilitate meaningful interpretation of the results obtained when deploying the tools of metric statistics and entail sensible inference, so that detected differences between groups are indeed relevant. In some spaces, feature preservation when transitioning from one data object to another along connecting geodesics that are determined by the metric is often desirable. This may include preservation of shape features, for example, unimodality in the case of distributions, which is a forte of the Wasserstein metric, or avoidance of the swelling effect in the case of symmetric positive definite matrices as provided, for example, by the Cholesky metric (Lin (2019)). A metric in the space of distributions that complements the given metric in the object space is utilized in the space of distance profiles, which correspond to one-dimensional distributions. In our approach, we use the 2-Wasserstein metric and exploit its connection with optimal transports; other metrics could be explored as well. An important property of the proposed distance profiles is that in metric spaces of (strong) negative type they characterize the underlying probability measure P, which guarantees that the proposed test attains asymptotic power against any alternative $Q \neq P$, and any alternative metric would need to match this property. Another obvious extension to consider is to use weighted transports in the definition of the distance profiles, where one could give more weight to the transported mass situated closer to $u = 0$.

8.2. *Outlook and future research.* Distance profiles and their metric lead to a new type of MDS for random objects, providing a representation of data objects that complements the more standard MDS representations based on the original metric in the object space, as exemplified in Section 7.3. The resulting visualization proved to be useful and interpretable in the examples we studied. But this is only a small start and visualization of random objects remains a widely open topic.

As we demonstrate in the numerical experiments in Section 6.3, in various scenarios the distance profile based test outperforms the energy test in terms of power. This is likely due to the fact that profile distances provide a more fine tuned assessment of the underlying measure P than means do, where energy tests are based on the latter. In our experiments, the profile based tests are less powerful than the energy test when the alternatives reflect mean shifts but more powerful for alternatives based on scale changes and also when heavy tails are involved. Further investigation for this phenomenon as well as potential improvements and modifications of the proposed test, for example, by judicious choice of the weights in (24), will be left for future research.

The proposed transport ranks can also be utilized to arrive at a new measure of object depth, complementing recent developments that extend classical notions of depth from Euclidean data to random objects (Cholaquidis, Fraiman and Moreno (2023), Dai and Lopez-Pintado (2023), Geenens, Nieto-Reyes and Francisci (2023)). Exploring these connections is a topic for future research. The properties of the proposed transport quantiles also deserve further study in view of the importance and challenges of defining quantiles in metric spaces. Specifically, rates and especially optimal rates of convergence for transport medians, transport quantiles and other tools of metric statistics will require future research efforts.

While we have utilized optimal transports of distance profiles to obtain transport ranks and transport quantiles, the concept of transports can be extended to objects in uniquely geodesic

spaces (Zhu and Müller (2023b)) and can then be used as a general modeling tool. Other recent developments include more sophisticated representations of random objects in reproducing kernel Hilbert spaces (Bhattacharjee, Li and Xue (2023)). These and similar developments are expected to provide valuable new tools for the nascent field of metric statistics. For random objects, essentially all relevant statistical methods for Euclidean data need to be redesigned with new theoretical justifications. Examples include deep learning that could be applied for Fréchet regression when the predictors are high-dimensional and principal component analysis for random objects, where at this point there is no general theoretically supported method. These are just a few of the many challenging open problems for future exploration.

Acknowledgments. We wish to thank two anonymous referees, an Associate Editor, and the Editor for their helpful and constructive comments which led to numerous improvements in the paper.

Data used in preparation of Section 7.5.2 in this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators and databases can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

This article is an expanded version of the Rietz Lecture delivered on June 27, 2022, by H.G.M. at the IMS Meeting in London and is an updated version of the paper posted at arXiv:2202.06117.

The first two authors contributed equally to the paper.

Funding. This research was supported in part by NSF Grants DMS-2311034 (PD), DMS-2311035 (YC), DMS-2014626 and DMS-2310450 (HGM).

SUPPLEMENTARY MATERIAL

Supplement to "Metric statistics: Exploration and inference for random objects with distance profiles" (DOI: 10.1214/24-AOS2368SUPP; .pdf). The Supplement contains proofs and auxiliary results, additional simulations for the two-sample test, additional simulations for distance profiles and transport ranks for multimodal multivariate data and distributional data as well as additional details for human mortality data and applications of distance profiles for Manhattan Yellow Taxi networks.

REFERENCES

AHIDAR-COUTRIX, A., LE GOUIC, T. and PARIS, Q. (2020). Convergence rates for empirical barycenters in metric spaces: Curvature, convexity and extendable geodesics. *Probab. Theory Related Fields* **177** 323–368. MR4095017 https://doi.org/10.1007/s00440-019-00950-0

AITCHISON, J. (1986). *The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability*. CRC Press, London. MR0865647 https://doi.org/10.1007/978-94-009-4109-0

AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed. *Lectures in Mathematics ETH Zürich*. Birkhäuser, Basel. MR2401600

BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634 https://doi.org/10.1126/science.286.5439.509

BARDEN, D., LE, H. and OWEN, M. (2018). Limiting behaviour of Fréchet means in the space of phylogenetic trees. *Ann. Inst. Statist. Math.* **70** 99–129. MR3742820 https://doi.org/10.1007/s10463-016-0582-9

BHATTACHARJEE, S., LI, B. and XUE, L. (2023). Nonlinear global Fréchet regression for random objects via weak conditional expectation. arXiv preprint. Available at arXiv:2310.07817.

BHATTACHARJEE, S. and MÜLLER, H.-G. (2023). Single index Fréchet regression. *Ann. Statist.* **51** 1770–1798. MR4658576 https://doi.org/10.1214/23-aos2307

BIGOT, J., GOUET, R., KLEIN, T. and LÓPEZ, A. (2017). Geodesic PCA in the Wasserstein space by convex PCA. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 1–26. MR3606732 https://doi.org/10.1214/15-AIHP706

BILLARD, L. and DIDAY, E. (2003). From the statistics of data to the statistics of knowledge: Symbolic data analysis. *J. Amer. Statist. Assoc.* **98** 470–487. MR1982575 https://doi.org/10.1198/016214503000242

BILLERA, L. J., HOLMES, S. P. and VOGTMANN, K. (2001). Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27** 733–767. MR1867931 https://doi.org/10.1006/aama.2001.0759

BISWAL, B., YETKIN, F. Z., HAUGHTON, V. M. and HYDE, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* **34** 537–541. https://doi.org/10.1002/mrm.1910340409

BLEI, R., GAO, F. and LI, W. V. (2007). Metric entropy of high dimensional distributions. *Proc. Amer. Math. Soc.* **135** 4009–4018. MR2341952 https://doi.org/10.1090/S0002-9939-07-08935-6

BOLSTAD, B. M., IRIZARRY, R. A., ÅSTRAND, M. and SPEED, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19** 185–193.

BURAGO, D., BURAGO, Y. and IVANOV, S. (2001). *A Course in Metric Geometry*. *Graduate Studies in Mathematics* **33**. Amer. Math. Soc., Providence, RI. MR1835418 https://doi.org/10.1090/gsm/033

CHAVEL, I. (2006). *Riemannian Geometry*: *A Modern Introduction*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **98**. Cambridge Univ. Press, Cambridge. MR2229062 https://doi.org/10.1017/CBO9780511616822

CHEN, H. and FRIEDMAN, J. H. (2017). A new graph-based two-sample test for multivariate and object data. *J. Amer. Statist. Assoc.* **112** 397–409. MR3646580 https://doi.org/10.1080/01621459.2016.1147356

CHEN, K., DELICADO, P. and MÜLLER, H.-G. (2017). Modelling function-valued stochastic processes, with applications to fertility dynamics. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 177–196. MR3597969 https://doi.org/10.1111/rssb.12160

CHEN, Y., DUBEY, P. and MÜLLER, H.-G. (2024). ODP: Exploration for random objects using distance profiles R package version 0.1.0. Available at https://github.com/yqgchen/ODP.

CHEN, Y., GAJARDO, A., FAN, J., ZHONG, Q., DUBEY, P., HAN, K., BHATTACHARJEE, S. and MÜLLER, H.-G. (2020). frechet: Statistical analysis for random objects and non-Euclidean data. R package version 0.2.0. Available at https://CRAN.R-project.org/package=frechet.

CHEN, H., and MÜLLER, H.-G. (2023). Sliced Wasserstein regression. arXiv preprint. Available at arXiv:2306.10601.

CHEN, Y., LIN, Z. and MÜLLER, H.-G. (2023). Wasserstein regression. *J. Amer. Statist. Assoc.* **118** 869–882. MR4595462 https://doi.org/10.1080/01621459.2021.1956937

CHEN, Y. and MÜLLER, H.-G. (2022). Uniform convergence of local Fréchet regression with applications to locating extrema and time warping for metric space valued trajectories. *Ann. Statist.* **50** 1573–1592. MR4441132 https://doi.org/10.1214/21-aos2163

CHENG, M.-Y. and WU, H.-T. (2013). Local linear regression on manifolds and its geometric interpretation. *J. Amer. Statist. Assoc.* **108** 1421–1434. MR3174718 https://doi.org/10.1080/01621459.2013.827984

CHOLAQUIDIS, A., FRAIMAN, R. and MORENO, L. (2023). Level sets of depth measures in abstract spaces. *TEST* **32** 942–957. MR4656906 https://doi.org/10.1007/s11749-023-00858-x

CORNEA, E., ZHU, H., KIM, P. and IBRAHIM, J. G. (2017). Regression models on Riemannian symmetric spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 463–482. MR3611755 https://doi.org/10.1111/rssb.12169

CUTURI, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems* 2292–2300.

DABO-NIANG, S. (2002). Estimation de la densité dans un espace de dimension infinie: Application aux diffusions. *C. R. Math. Acad. Sci. Paris* **334** 213–216. MR1891061 https://doi.org/10.1016/S1631-073X(02)02247-1

DAI, X. (2022). Statistical inference on the Hilbert sphere with application to random densities. *Electron. J. Stat.* **16** 700–736. MR4366819 https://doi.org/10.1214/21-ejs1942

DAI, X., LIN, Z. and MÜLLER, H.-G. (2021). Modeling sparse longitudinal data on Riemannian manifolds. *Biometrics* **77** 1328–1341. MR4357841 https://doi.org/10.1111/biom.13385

DAI, X. and LOPEZ-PINTADO, S. (2023). Tukey's depth for object data. *J. Amer. Statist. Assoc.* **118** 1760–1772. Authors writing for the Alzheimer's Disease Neuroimaging Initiative. MR4646604 https://doi.org/10.1080/01621459.2021.2011298

DONG, Y. and WU, Y. (2022). Fréchet kernel sliced inverse regression. *J. Multivariate Anal.* **191** Paper No. 105032, 14. MR4432318 https://doi.org/10.1016/j.jmva.2022.105032

DRYDEN, I. L., KOLOYDENKO, A. and ZHOU, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.* **3** 1102–1123. MR2750388 https://doi.org/10.1214/09-AOAS249

DRYDEN, I. L. and MARDIA, K. V. (2016). *Statistical Shape Analysis with Applications in R*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3559734 https://doi.org/10.1002/9781119072492

DUBEY, P., CHEN, Y. and MÜLLER, H.-G. (2024). Supplement to "Metric statistics: Exploration and inference for random objects With distance profiles." https://doi.org/10.1214/24-AOS2368SUPP

DUBEY, P. and MÜLLER, H.-G. (2019). Fréchet analysis of variance for random objects. *Biometrika* **106** 803–821. MR4031200 https://doi.org/10.1093/biomet/asz052

DUBEY, P. and MÜLLER, H.-G. (2020a). Functional models for time-varying random objects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 275–327. MR4084166

DUBEY, P. and MÜLLER, H.-G. (2020b). Fréchet change-point detection. *Ann. Statist.* **48** 3312–3335. MR4185810 https://doi.org/10.1214/19-AOS1930

ELTZNER, B. and HUCKEMANN, S. F. (2019). A smeary central limit theorem for manifolds with application to high-dimensional spheres. *Ann. Statist.* **47** 3360–3381. MR4025745 https://doi.org/10.1214/18-AOS1781

FERAGEN, A., LAUZE, F. and HAUBERG, S. (2015). Geodesic exponential kernels: When curvature and linearity conflict. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 3032–3042.

FILZMOSER, P., HRON, K. and TEMPL, M. (2019). *Applied Compositional Data Analysis*: *With Worked Examples in R*. Springer.

FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. Henri Poincaré* **10** 215–310. MR0027464

FRISTON, K. J., FRITH, C. D., LIDDLE, P. F. and FRACKOWIAK, R. S. J. (1993). Functional connectivity: The principal-component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.* **13** 5–14.

GAO, F. and WELLNER, J. A. (2009). On the rate of convergence of the maximum likelihood estimator of a k-monotone density. *Sci. China Ser. A* **52** 1525–1538. MR2520591 https://doi.org/10.1007/s11425-009-0102-y

GARBA, M. K., NYE, T. M. W., LUEG, J. and HUCKEMANN, S. F. (2021). Information geometry for phylogenetic trees. *J. Math. Biol.* **82** Paper No. 19, 39. MR4218000 https://doi.org/10.1007/s00285-021-01553-x

GEENENS, G., NIETO-REYES, A. and FRANCISCI, G. (2023). Statistical depth in abstract metric spaces. *Stat. Comput.* **33** Paper No. 46, 15. MR4554146 https://doi.org/10.1007/s11222-023-10216-4

GHODRATI, L. and PANARETOS, V. M. (2023). On distributional autoregression and iterated transportation. arXiv preprint. Available at arXiv:2303.09469.

GHOSAL, A., MEIRING, W. and PETERSEN, A. (2023). Fréchet single index models for object response regression. *Electron. J. Stat.* **17** 1074–1112. MR4575027 https://doi.org/10.1214/23-ejs2120

GHOSAL, R., VARMA, V. R., VOLFSON, D., HILLEL, I., URBANEK, J., HAUSDORFF, J. M., WATTS, A. and ZIPUNNIKOV, V. (2023). Distributional data analysis via quantile functions and its application to modeling digital biomarkers of gait in Alzheimer's Disease. *Biostatistics* **24** 539–561. MR4615240 https://doi.org/10.1093/biostatistics/kxab041

GINESTET, C. E., LI, J., BALACHANDRAN, P., ROSENBERG, S. and KOLACZYK, E. D. (2017). Hypothesis testing for network data in functional neuroimaging. *Ann. Appl. Stat.* **11** 725–750. MR3693544 https://doi.org/10.1214/16-AOAS1015

HRON, K., MENAFOGLIO, A., TEMPL, M., HRŮZOVÁ, K. and FILZMOSER, P. (2016). Simplicial principal component analysis for density functions in Bayes spaces. *Comput. Statist. Data Anal.* **94** 330–350. MR3412829 https://doi.org/10.1016/j.csda.2015.07.007

HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis*, *with an Introduction to Linear Operators*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3379106 https://doi.org/10.1002/9781118762547

HUCKEMANN, S. F. and ELTZNER, B. (2021). Data analysis on nonstandard spaces. *Wiley Interdiscip. Rev.*: *Comput. Stat.* **13** Paper No. e1526, 19. MR4242812 https://doi.org/10.1002/wics.1526

JEON, J. M. and PARK, B. U. (2020). Additive regression with Hilbertian responses. *Ann. Statist.* **48** 2671–2697. MR4152117 https://doi.org/10.1214/19-AOS1902

JUNG, S., DRYDEN, I. L. and MARRON, J. S. (2012). Analysis of principal nested spheres. *Biometrika* **99** 551–568. MR2966769 https://doi.org/10.1093/biomet/ass022

JUNG, S., SCHWARTZMAN, A. and GROISSER, D. (2015). Scaling-rotation distance and interpolation of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **36** 1180–1201. MR3379023 https://doi.org/10.1137/140967040

KANTOROVITCH, L. (1958). On the translocation of masses. *Manage. Sci.* **5** 1–4. MR0096552 https://doi.org/10.1287/mnsc.5.1.1

KIM, J., ROSENBERG, N. A. and PALACIOS, J. A. (2020). Distance metrics for ranked evolutionary trees. *Proc. Natl. Acad. Sci. USA* **117** 28876–28886.

KLEBANOV, L. B. (2005). *N-Distances and Their Applications*. Karolinum Press, Charles Univ. Prague, Czech Republic.

KNEIP, A. and UTIKAL, K. J. (2001). Inference for density families using functional principal component analysis. *J. Amer. Statist. Assoc.* **96** 519–542. With comments and a rejoinder by the authors. MR1946423 https://doi.org/10.1198/016214501753168235

KOLACZYK, E. D., LIN, L., ROSENBERG, S., WALTERS, J. and XU, J. (2020). Averages of unlabeled networks: Geometric characterization and asymptotic behavior. *Ann. Statist.* **48** 514–538. MR4065172 https://doi.org/10.1214/19-AOS1820

KOLOURI, S., NADJAHI, K., SIMSEKLI, U., BADEAU, R. and ROHDE, G. (2019). Generalized sliced Wasserstein distances. *Adv. Neural Inf. Process. Syst.* **32** 261–272.

KOLOURI, S., ZOU, Y. and ROHDE, G. K. (2016). Sliced Wasserstein kernels for probability distributions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 5258–5267.

LIN, Z. (2019). Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition. *SIAM J. Matrix Anal. Appl.* **40** 1353–1370. MR4032859 https://doi.org/10.1137/18M1221084

LIN, Z. and MÜLLER, H.-G. (2021). Total variation regularized Fréchet regression for metric-space valued data. *Ann. Statist.* **49** 3510–3533. MR4352539 https://doi.org/10.1214/21-aos2095

LINDQUIST, M. A. (2008). The statistical analysis of fMRI data. *Statist. Sci.* **23** 439–464. MR2530545 https://doi.org/10.1214/09-STS282

LIU, R. Y. and SINGH, K. (1992). Ordering directional data: Concepts of data depth on circles and spheres. *Ann. Statist.* **20** 1468–1484. MR1186260 https://doi.org/10.1214/aos/1176348779

LUEG, J., GARBA, M. K., NYE, T. M. W. and HUCKEMANN, S. F. (2022). Foundations of the Wald space for phylogenetic trees. arXiv preprint. Available ar arXiv:2209.05332.

LUNAGÓMEZ, S., OLHEDE, S. C. and WOLFE, P. J. (2021). Modeling network populations via graph distances. *J. Amer. Statist. Assoc.* **116** 2023–2040. MR4353730 https://doi.org/10.1080/01621459.2020.1763803

LYONS, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* **41** 3284–3305. MR3127883 https://doi.org/10.1214/12-AOP803

MARDIA, K. V. (1978). Some properties of classical multi-dimensional scaling. *Comm. Statist. Theory Methods* **7** 1233–1241. MR0514645 https://doi.org/10.1080/03610927808827707

MARRON, J. S. and DRYDEN, I. L. (2021). *Object Oriented Data Analysis*. CRC Press, Boca Raton.

MATABUENA, M., PETERSEN, A., VIDAL, J. C. and GUDE, F. (2021). Glucodensities: A new representation of glucose profiles using distributional data analysis. *Stat. Methods Med. Res.* **30** 1445–1464. MR4269959 https://doi.org/10.1177/0962280221998064

MÜLLER, H.-G. (2016). Peter Hall, functional data analysis and random objects. *Ann. Statist.* **44** 1867–1887. MR3546436 https://doi.org/10.1214/16-AOS1492

PANARETOS, V. M. and ZEMEL, Y. (2020). *An Invitation to Statistics in Wasserstein Space*. *SpringerBriefs in Probability and Mathematical Statistics*. Springer, Cham. MR4350694 https://doi.org/10.1007/978-3-030-38438-8

PEGORARO, M. and BERAHA, M. (2022). Projected statistical methods for distributional data on the real line with the Wasserstein metric. *J. Mach. Learn. Res.* **23** Paper No. [37], 59. MR4420762

PETERSEN, A. and MÜLLER, H.-G. (2016a). Functional data analysis for density functions by transformation to a Hilbert space. *Ann. Statist.* **44** 183–218. MR3449766 https://doi.org/10.1214/15-AOS1363

PETERSEN, A. and MÜLLER, H.-G. (2016b). Fréchet integration and adaptive metric selection for interpretable covariances of multivariate functional data. *Biometrika* **103** 103–120. MR3465824 https://doi.org/10.1093/biomet/asv054

PETERSEN, A. and MÜLLER, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719. MR3909947 https://doi.org/10.1214/17-AOS1624

PETERSEN, A., ZHANG, C. and KOKOSZKA, P. (2022). Modeling probability density functions as data objects. *Econom. Stat.* **21** 159–178. MR4366852 https://doi.org/10.1016/j.ecosta.2021.04.004

PIGOLI, D., ASTON, J. A. D., DRYDEN, I. L. and SECCHI, P. (2014). Distances and inference for covariance operators. *Biometrika* **101** 409–422. MR3215356 https://doi.org/10.1093/biomet/asu008

POWER, J. D., COHEN, A. L., NELSON, S. M., WIG, G. S., BARNES, K. A., CHURCH, J. A., VOGEL, A. C., LAUMANN, T. O., MIEZIN, F. M. et al. (2011). Functional network organization of the human brain. *Neuron* **72** 665–678. https://doi.org/10.1016/j.neuron.2011.09.006

R CORE TEAM (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

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

SCEALY, J. L. and WELSH, A. H. (2014). Colours and cocktails: Compositional data analysis 2013 Lancaster lecture. *Aust. N. Z. J. Stat.* **56** 145–169. MR3226434 https://doi.org/10.1111/anzs.12073

SCHOENBERG, I. J. (1937). On certain metric spaces arising from Euclidean spaces by a change of metric and their imbedding in Hilbert space. *Ann. of Math.* (2) **38** 787–793. MR1503370 https://doi.org/10.2307/1968835

SCHOENBERG, I. J. (1938). Metric spaces and positive definite functions. *Trans. Amer. Math. Soc.* **44** 522–536. MR1501980 https://doi.org/10.2307/1989894

SCHÖTZ, C. (2019). Convergence rates for the generalized Fréchet mean via the quadruple inequality. *Electron. J. Stat.* **13** 4280–4345. MR4023955 https://doi.org/10.1214/19-EJS1618

SCHÖTZ, C. (2022). Nonparametric regression in nonstandard spaces. *Electron. J. Stat.* **16** 4679–4741. MR4489238 https://doi.org/10.1214/22-ejs2056

SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291. MR3127866 https://doi.org/10.1214/13-AOS1140

SEVERN, K. E., DRYDEN, I. L. and PRESTON, S. P. (2022). Manifold valued data analysis of samples of networks, with applications in corpus linguistics. *Ann. Appl. Stat.* **16** 368–390. MR4400514 https://doi.org/10.1214/21-aoas1480

STEINKE, F. and HEIN, M. (2009). Non-parametric regression between manifolds. *Adv. Neural Inf. Process. Syst.* 1561–1568.

STEINKE, F., HEIN, M. and SCHÖLKOPF, B. (2010). Nonparametric regression between general Riemannian manifolds. *SIAM J. Imaging Sci.* **3** 527–563. MR2736019 https://doi.org/10.1137/080744189

STURM, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. In *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces* (*Paris*, 2002). *Contemp. Math.* **338** 357–390. Amer. Math. Soc., Providence, RI. MR2039961 https://doi.org/10.1090/conm/338/06080

SZÉKELY, G. J. and RIZZO, M. L. (2004). Testing for equal distributions in high dimension. *Interstate* **5** 1–6.

SZÉKELY, G. J. and RIZZO, M. L. (2017). The energy of data. *Annu. Rev. Stat. Appl.* **4** 447–479.

TUCKER, D. C., WU, Y. and MÜLLER, H.-G. (2023). Variable selection for global Fréchet regression. *J. Amer. Statist. Assoc.* **118** 1023–1037. MR4595474 https://doi.org/10.1080/01621459.2021.1969240

VAKHANIA, N. N., TARIELADZE, V. I. and CHOBANYAN, S. A. (1987). *Probability Distributions on Banach Spaces. Mathematics and Its Applications* (*Soviet Series*) **14**. Reidel, Dordrecht. Translated from the Russian and with a preface by Wojbor A. Woyczynski. MR1435288 https://doi.org/10.1007/978-94-009-3873-1

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*: *With applications to statistics. Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

VENET, N. (2019). Nonexistence of fractional Brownian fields indexed by cylinders. *Electron. J. Probab.* **24** Paper No. 75, 26. MR3978225 https://doi.org/10.1214/18-EJP256

VIRTA, J., LEE, K.-Y. and LI, L. (2022). Sliced inverse regression in metric spaces. *Statist. Sinica* **32** 2315–2337. MR4485085

WANG, H. and MARRON, J. S. (2007). Object oriented data analysis: Sets of trees. *Ann. Statist.* **35** 1849–1873. MR2363955 https://doi.org/10.1214/009053607000000217

WANG, J.-L., CHIOU, J.-M. and MÜLLER, H.-G. (2016). Functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295.

WANG, X., ZHU, J., PAN, W., ZHU, J. and ZHANG, H. (2023). Nonparametric statistical inference via metric distribution function in metric spaces. *J. Amer. Statist. Assoc.* (to appear). https://doi.org/10.1080/01621459.2023.2277417.

YUAN, Y., ZHU, H., LIN, W. and MARRON, J. S. (2012). Local polynomial regression for symmetric positive definite matrices. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 697–719. MR2965956 https://doi.org/10.1111/j.1467-9868.2011.01022.x

ZEMEL, Y. and PANARETOS, V. M. (2019). Fréchet means and Procrustes analysis in Wasserstein space. *Bernoulli* **25** 932–976. MR3920362 https://doi.org/10.3150/17-bej1009

ZHANG, Q., LI, B. and XUE, L. (2024). Nonlinear sufficient dimension reduction for distribution-on-distribution regression. *J. Multivariate Anal.* **202** Paper No. 105302. MR4711112 https://doi.org/10.1016/j.jmva.2024.105302

ZHANG, Q., XUE, L. and LI, B. (2021). Dimension reduction and data visualization for Fréchet regression. arXiv preprint. Available at arXiv:2110.00467.

ZHOU, H. and MÜLLER, H.-G. (2023). Optimal transport representations and functional principal components for distribution-valued processes. arXiv preprint. Available at arXiv:2310.20088.

ZHOU, Y. and MÜLLER, H.-G. (2022). Network regression with graph Laplacians. *J. Mach. Learn. Res.* **23** Paper No. [320], 41. MR4577759 https://doi.org/10.22405/2226-8383-2022-23-5-320-336

ZHU, C. and MÜLLER, H.-G. (2023a). Autoregressive optimal transport models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 1012–1033.

ZHU, C. and MÜLLER, H.-G. (2023b). Geodesic optimal transport regression. arXiv preprint. Available at arXiv:2312.15376.

ZHU, C. and MÜLLER, H.-G. (2024). Spherical autoregressive models, with application to distributional and compositional time series. *J. Econometrics* **239** Paper No. 105389, 16. MR4708615 https://doi.org/10.1016/j.jeconom.2022.12.008