

# GEODESIC MIXED EFFECTS MODELS FOR REPEATEDLY OBSERVED/LONGITUDINAL RANDOM OBJECTS

Satarupa Bhattacharjee<sup>1</sup> and Hans-Georg Müller<sup>2</sup>

<sup>1</sup> Department of Statistics, Pennsylvania State University

<sup>2</sup>Department of Statistics, University of California, Davis

## Abstract

Mixed effect modeling for longitudinal data is challenging when the observed data are random objects, which are complex data taking values in a general metric space without linear structure. In such settings the classical additive error model and distributional assumptions are unattainable. Due to the rapid advancement of technology, longitudinal data containing complex random objects, such as covariance matrices, data on Riemannian manifolds, and probability distributions are becoming more common. Addressing this challenge, we develop a mixed-effects regression for data in geodesic spaces, where the underlying mean response trajectories are geodesics in the metric space and the deviations of the observations from the model are quantified by perturbation-maps or transports. A key finding is that the geodesic trajectories assumption for the case of random objects is a natural extension of the linearity assumption in the standard Euclidean scenario. Further, geodesics can be recovered from noisy observations by exploiting a connection between the geodesic path and the path obtained by global Fréchet regression for random objects. The effect of baseline Euclidean covariates on the geodesic paths is modeled by another Fréchet regression step. We study the asymptotic convergence of the proposed estimates and provide illustrations through simulations and real-data applications.

KEY WORDS: Random Effects; Random objects; Geodesics; Perturbation; Optimal transport; Fréchet regression; M-estimation.

arXiv:2307.05726v1 [stat.ME] 11 Jul 2023

# 1 Introduction

In the era of modern data science, complex data structures are increasingly encountered. An important but largely unexplored setting is where a response variable takes values in a non-Euclidean metric space without vector space operations or inner product. Examples of such random objects (Müller, 2016) include distributional data in Wasserstein space (Petersen and Müller, 2016; Matabuena et al., 2021), symmetric positive definite matrix objects (Dryden et al., 2009), spherical data (Di Marzio et al., 2014), phylogenetic trees (Billera et al., 2001) and data on finite-dimensional Riemannian manifolds (Bhattacharya and Patrangenaru, 2003, 2005; Afsari, 2011; Eltzner and Huckemann, 2019), among other data types. Data modeling and analysis for metric space valued data is challenging due to the absence of any linear structure. For example, the definition of a sample or population mean as an average or expected value is not applicable and is replaced by barycenters or Fréchet means (Fréchet, 1948). Similarly, regression approaches to quantify the dependence between a random object response and Euclidean predictors require a notion of a conditional Fréchet mean (Petersen and Müller, 2019) with several approaches for corresponding regression models (Hein, 2009; Dong and Wu, 2022; Schötz, 2022; Zhang et al., 2021, 2022).

Technological advances have made it possible to record and efficiently store repeated measurements of images (Peyré, 2009; González-Briones et al., 2018), shapes (Small, 2012), networks (Tsochantaridis et al., 2004) and other random objects. There are only few methods available to analyze time courses of random objects and only for the case where time courses are continuously recorded and fully observed over time (Dubey and Müller, 2020). But when such data are recorded in longitudinal studies with repeated observations of random objects, these are often sparsely recorded over time, posing a substantial additional challenge for statistical analysis. To our knowledge, there is currently no statistical method available to handle longitudinal random objects. This paper presents the first approach for the statistical analysis of such data. For sparsely sampled trajectories as we consider here it is of interest to gain information about the actual individual time courses, i.e., the underlying metric-space valued curves that produce the observed measurements but are latent, due to the sparse measurement scheme.

Flexible nonparametric recovery methods have been extensively studied for the case of scalar responses based on versions of functional principal component analysis (see, e.g., Staniswalis and

Lee, 1998; Rice and Wu, 2001; Yao et al., 2005; Şentürk and Nguyen, 2011; Yao et al., 2015; Chen et al., 2021; Li et al., 2022). However, all of these approaches require that the data are in a linear space and thus cannot be extended to the case of object data, where one cannot make use of vector space operations. A second and more restrictive approach are classical Euclidean linear mixed effects models (Laird and Ware, 1982; Diggle et al., 2002; Verbeke et al., 2010), where the individuals in the population are assumed to follow the same general linear model but with random intercepts and slopes that are subject-specific, with various extensions (see, e.g., Wu, 2009; Schiratti et al., 2015; Allasonniere et al., 2017; Yue et al., 2020; Pellagatti et al., 2021). Our goal in this paper is to address the challenges to extend random effects models to the case of object data.

Given a covariate vector  $Z_i \in \mathbb{R}^p$ ,  $p \geq 1$ , for the  $i^{\text{th}}$  subject,  $1 \leq i \leq n$ , repeated measurements  $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})$  and measurement times  $T_i = (T_{i1}, T_{i2}, \dots, T_{in_i})$ , the mixed effects linear regression for repeated measurements/longitudinal data is

$$\mathbb{E}(Y_i(t)|\nu_i, T_{ij} = t) = \nu_i t, \quad \mathbb{E}(\nu_i|Z_i = z) = \beta^\top z, \quad (1)$$

where the  $\nu_i$  are subject-specific random slopes that determine trajectories  $\nu_i t$  and depend linearly on the baseline covariate vector  $Z$ . Here  $\beta \in \mathbb{R}^p$  is a fixed parameter vector. A typical additional assumption is  $Y_i(t) = \nu_i t + \varepsilon(t)$  for zero mean finite variance additive errors and also joint Gaussianity of all random components. As we aim to generalize model (1) to the case of sparse random object observations  $Y_{ij}$ , where an additive structure for the model is not available, the trajectories  $\nu_i t$  are written without intercepts; in the real case, this form can be obtained by centering predictors and responses for each subject.

A key observation that makes it possible to generalize model (1) to the case of object data is that the linearity assumption from a more general perspective corresponds to the assumption that responses are scattered around a geodesic, which in the case of real-valued data is a line. Accordingly we consider in the following geodesic metric spaces; we will model subject-specific random trajectories as geodesics in such spaces. Noisy observations of random objects are sparse in time and located around the geodesic, where noise is modeled through perturbation maps that are applied to the true random objects, as in metric spaces there is no framework for additive noise. To obtain asymptotic results, we consider the case of small errors and develop an approach that makes

it possible to recover the subject-specific geodesic trajectories, using global Fréchet regression for random object responses (Petersen and Müller, 2019) as an auxiliary tool.

In Section 2 we provide a brief review of metric geometry and geodesics and provide further motivation for the proposed model. In Section 3, we discuss the connection between the underlying subject-specific geodesic path and the path estimated by the global Fréchet regression method and proceed to establish theoretical guarantees for the asymptotic convergence of model components, including rates, based on M-estimation theory. Our motivating application examples deal with samples of probability distributions, data lying on the unit sphere in  $\mathbb{R}^3$  and correlation matrices, which are illustrated with simulations in Section 4. Real data applications for resting state fMRI longitudinal data from ADNI and demographic data are discussed in Section 5.

## 2 Preliminaries and Model

### 2.1 Preliminaries on Metric Spaces

In the following,  $(\mathcal{M}, d)$  denotes a metric space that is complete, separable and totally bounded and we refer to the elements  $Y \in \mathcal{M}$  as random objects. We consider sets  $\mathcal{T} = [0, 1]$  and  $\mathcal{S} \subset \mathbb{R}^p$  for  $p \geq 1$  and a random tuple  $(Y, T, Z)$  with a joint distribution on the product space  $\mathcal{M} \times \mathcal{T} \times \mathcal{S}$ , where in a regression setting  $Y \in \mathcal{M}$  is a random object response,  $T \in \mathcal{T}$  is a random time point where the random object  $Y$  is observed and  $Z$  a (baseline) covariate with  $Z \in \mathcal{S}$ . We focus on a longitudinal setting, where one observes  $n$  subjects and  $n_i > 1$  observations are made at random times  $T_{ij} \in \mathcal{T}$  for the  $i^{\text{th}}$  subject with corresponding observations  $Y_{ij} = Y_i(T_{ij}) \in \mathcal{M}$ .

A geodesic in a geodesic metric space connecting two distinct points is the shortest path connecting the two points. Geodesics in a metric space are analogous to straight lines in a Euclidean space. In a uniquely geodesic metric space  $\mathcal{M}$  with metric  $d$ , a constant speed geodesic  $\gamma_{\nu_0, \nu_1}(t) \in \mathcal{M}$ ,  $t \in [0, 1]$ , connecting two points  $\nu_0$  and  $\nu_1$  is characterized by  $\gamma_{\nu_0, \nu_1}(0) = \nu_0$ ,  $\gamma_{\nu_0, \nu_1}(1) = \nu_1$  and  $d(\gamma_{\nu_0, \nu_1}(t_1), \gamma_{\nu_0, \nu_1}(t_2)) = |t_1 - t_2|d(\nu_0, \nu_1)$ . If for any two points in a metric space there exists a geodesic that connects them, the space is a geodesic space and it is uniquely geodesic if for every pair of points  $x, y \in \mathcal{M}$ , there is a unique geodesic  $\gamma_{\nu_0, \nu_1} : [0, 1] \mapsto \mathcal{M}$  from  $x$  to  $y$ . For further details and background we refer to Burago et al. (2001) and the review in Section

2 of Lin and Müller (2021). Given a geodesic  $\gamma_{\nu_0, \nu_1}(t)$  defined on  $t \in [0, 1]$ , if the geodesic property as defined above continues to hold for  $\gamma_{\nu_0, \nu_1}(t)$  with  $t \in [t_1, t_2]$  where  $t_1 < 0 < 1 < t_2$ , we say that the geodesic can be extended from  $[0, 1]$  to  $[t_1, t_2]$  (Ahidar-Coutrix et al., 2020). We assume throughout that  $(\mathcal{M}, d)$  is a uniquely extendable geodesic space, i.e., it is a uniquely geodesic space, where all geodesics can be extended. It is obvious that the Euclidean space, where the geodesic path connecting two points  $a, b \in \mathbb{R}$  is simply the line connecting the two points, is a uniquely extendable geodesic space. Other examples of uniquely extendable geodesic spaces are as follows.

*Example 1: Space of distributions with the Wasserstein metric.* For a closed interval  $Q \subset \mathbb{R}$ , the Wasserstein space  $\mathcal{W}_2(Q)$  of probability distributions on  $Q$  with finite second moments is endowed with the  $L_2$ -Wasserstein distance

$$d_W(\mu, \nu) = \left( \int_0^1 [F_\mu^{-1}(s) - F_\nu^{-1}(s)]^2 ds \right)^{1/2}, \text{ for } \mu, \nu \in \mathcal{W}_2(Q),$$

where  $F_\mu^{-1}$  and  $F_\nu^{-1}$  denote the quantile functions of  $\mu$  and  $\nu$ , respectively. We further require the distributions to be continuous, i.e., to possess densities. Then  $(\mathcal{W}_2(Q), d_W)$  is a uniquely geodesic space (Ambrosio and Gigli, 2008). Given any  $\mu, \nu \in \mathcal{W}_2(Q)$  where  $\mu \neq \nu$ , there is a unique geodesic that connects  $\mu$  and  $\nu$ , given by  $\gamma_{\mu, \nu}(t) = [t(F_\nu^{-1} \circ F_\mu - \text{id}) + \text{id}] \# \mu$ ,  $t \in [0, 1]$ . For a measurable function  $h : Q \rightarrow Q$ ,  $h \# \mu$  is a pushforward measure such that  $h \# \mu(A) = \mu(\{r \in Q : h(r) \in A\})$  for any set  $A \in \mathcal{B}(Q)$ , the Borel  $\sigma$ -algebra on  $Q$ . For the extendibility of geodesics in the space of continuous probability measures we refer to Ahidar-Coutrix et al. (2020); Zhu and Müller (2023).

*Example 2: Space of positive definite matrices.* The space of positive definite symmetric  $K \times K$  matrices  $\mathcal{S}_K$ , equipped with the Frobenius inner product  $\langle A, B \rangle_F = \text{tr}(A^\top B)$  and the induced Frobenius metric  $d_F(A, B) = \|A - B\|_F$ ,  $A, B \in \mathcal{S}_K$ , where  $\|A\|_F$  is the usual Euclidean matrix norm, possesses unique geodesics, which are straight lines in the Euclidean vector space given by  $\gamma_{A, B} : [0, 1] \rightarrow \mathcal{S}_K$  with  $\gamma_{A, B}(t) = tA + (1-t)B$ . Other metrics  $d$  for which  $\mathcal{S}_K$  is a uniquely geodesic space include the log-Euclidean metric (Arsigny et al., 2007), the power metric family (Dryden et al., 2010), the Log-Cholesky metric (Lin, 2019) and the Bures-Wasserstein metric (Takatsu, 2011); these geodesics are extendible as long as  $A, B$  are strictly positive definite. A popular metric on  $\mathcal{S}_K$  that has been successfully used in various practical applications for covariances is the square root power metric (Pigoli et al., 2014; Tavakoli et al., 2019), where  $d_{1/2}(A, B) = \|A^{1/2} - B^{1/2}\|_F$ ,  $A, B \in \mathcal{S}_K$ ; we

will use this metric in Section 5 to illustrate the proposed random effects model for neuroimaging data. The geodesics in this metric are  $\gamma_{A,B}(t) = (tA^{1/2} + (1-t)B^{1/2})^2$ .

*Example 3: Spheres with geodesic metric.* A  $(p-1)$ -dimensional sphere  $S^{p-1} = \{x \in \mathbb{R}^p : \|x\| = 1\}$  embedded in  $\mathbb{R}^p$  is a complete Riemannian manifold. The geodesic metric  $d_g$  between two points  $x, y$  on the surface of the unit sphere  $S^n$  is given by  $d_g(x, y) = \arccos\langle x, y \rangle$ . Consider  $M = S^2$  the 2-sphere with the spherical geodesic metric. Then the great circles are geodesics. The great circle passing through two points  $x, y \in S^2$  can be parametrized as  $\gamma_{u,v}(t) = (\cos t)u + (\sin t)v$ . However, this space is not uniquely geodesic as two polar points can be connected by arbitrarily many different geodesics. In order to make the space a uniquely geodesic space one can slice off the subset of the sphere with  $x_1 \leq -1 + \gamma$  for any small  $0 < \gamma \leq 1/2$ , which includes the half sphere, where  $x_1$  is the first coordinate of  $x$ . Since the sphere with the slice removed is an open set, the great circle geodesics are extendable.

*Example 4: The space of phylogenetic trees.* Phylogenetic trees are of interest in evolutionary biology, where they are used to represent the evolutionary history of a set of organisms. In a seminal paper (Billera et al., 2001), phylogenetic trees with  $m$  leaves are modeled by metric  $m$ -trees endowed with a metric that turns the space of phylogenetic  $m$ -trees into a metric space, as follows: A leaf is a vertex that is connected by only one edge, and a metric  $m$ -tree is a tree with  $m$  uniquely labeled leaves and positive lengths on all interior edges, where an edge is called an interior edge if it does not connect to a leaf. A collection of  $m$ -trees that have the same tree structure (taking leaf labels into account) but different edge lengths can be identified with the orthant  $(0, \infty)^r$ , where  $r$  is determined by the tree structure and corresponds to the number of interior edges of each tree in the collection. With this identification between points and metric  $m$ -trees, the BHV metric  $d_T$  on the space  $\mathcal{T}_m$  of all metric  $m$ -trees is defined as follows: For two trees in the same orthant, their distance is the Euclidean distance of their edge lengths, while for two trees from different orthants, their distance is the minimum length over all paths that connect them and consist of only connected segments, where a segment is a straight line within an orthant. The minimum length path is the geodesic, which is extendable within the orthants where it starts and ends. According to Lemma 4.1 of Billera et al. (2001),  $\mathcal{T}_m$  is a unique geodesic space. It is a CAT(0) space. More generally, each geodesic CAT(0) metric space is a unique geodesic space (for a brief review see, e.g., Lin and Müller, 2021).

## 2.2 Preliminaries on noisy trajectories

Since the metric space where the random object responses reside is devoid of any vector-space structure, one cannot use classical additive error models. Noise in observations can instead be quantified by perturbation maps (Chen and Müller, 2022)  $\mathcal{P} : \mathcal{M} \rightarrow \mathcal{M}$ , characterized by

$$\mu' = \operatorname{argmin}_{\mu \in \mathcal{M}} \mathbb{E}[d^2(\mathcal{P}(\mu'), \mu)] \text{ for all } \mu' \in \mathcal{M}. \quad (2)$$

We assume that for the  $i^{\text{th}}$  individual, noise-contaminated random objects  $Y_{ij}$  recorded at  $T_{ij}$  are centered around an underlying trajectory  $\alpha_i$ . With perturbation maps (2), the observed data are

$$Y_{ij} = \mathcal{P}_{ij}(\alpha_i(T_{ij})), \quad j = 1, \dots, n_i, \quad i = 1, \dots, n. \quad (3)$$

In connection with the classical mixed effects model in (1), the perturbation map replaces additive errors and the underlying trajectory is  $\alpha_i(t) = \nu_i t$ . The size of the error is quantified as  $\mathbb{E}[d^2(\mathcal{P}_{ij}(\alpha_i(t)), \alpha_i(t))]$ , which is bounded owing to the total boundedness of the metric space, and corresponds to the error variance for classical Euclidean responses.

For the classical linear mixed model  $\alpha_i(t) = \nu_i t$  is a line in the Euclidean space and therefore a geodesic. Thus a defining feature of the classical linear mixed effects model is to fit geodesics to the data. A natural extension to the case of a general geodesic space is then to replace linearity by geodesicity, where observed data are assumed to cluster around a true geodesic. For the remainder of the paper, the underlying trajectory  $\alpha_i$  for the  $i^{\text{th}}$  individual is assumed to be a uniquely extendable geodesic  $\alpha_i = \gamma_{\nu_{i0}, \nu_{i1}}^{(i)}$  in the metric space  $(\mathcal{M}, d)$  connecting the points  $\nu_{i0}$  and  $\nu_{i1}$ . This leads to the following general model for the observed data,

$$Y_{ij} = \mathcal{P}_{ij}[\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(T_{ij})], \quad j = 1, \dots, n_i, \quad i = 1, \dots, n. \quad (4)$$

## 2.3 Random effects model for $\mathcal{M}$ -valued data

In a uniquely geodesic space  $\mathcal{M}$  the randomness of the geodesic path  $\gamma_{\nu_0, \nu_1}(\cdot)$  is incorporated through the two endpoints  $\nu_0$  and  $\nu_1$  that determine the geodesic. For the  $i^{\text{th}}$  individual, the underlying true geodesic path that connects the end-points  $\nu_{i0}$  and  $\nu_{i1}$  is  $\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(t) : [0, 1] \rightarrow (\mathcal{M}, d)$ . We assume

throughout that with probability 1 the random geodesic that generates the observations is unique, an assumption that is satisfied for unique geodesic spaces such as those discussed in Examples 1-4 in Section 2.1. We also require the following assumption for the data generation mechanism.

- (A1) Observation times  $T_{ij}$ , random perturbation maps  $\mathcal{P}_{ij}$  and the random mechanism that generates the underlying geodesic trajectory  $\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(t) \ t \in [0, 1]$  (or alternatively generates the two endpoints  $\nu_{i0}$  and  $\nu_{i1}$ ) are all independent and i.i.d.

The proposed random effects model at the subject level for  $\mathcal{M}$ -valued responses is

$$\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(T_{ij}) = \underset{\mu \in \mathcal{M}}{\operatorname{argmin}} \mathbb{E}[d^2(Y_{ij}, \mu) | \gamma_{\nu_{i0}, \nu_{i1}}^{(i)}, T_{ij}], \quad Y_{ij} = \mathcal{P}_{ij}[\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(T_{ij})], \quad j = 1, \dots, n_i. \quad (5)$$

Once the random effects inherent in the subject-specific geodesics are recovered from the noisy observations, we regress the entire geodesic paths  $\{\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(t) : t \in [0, 1]\}$  that constitute the responses on the predictors  $Z_i \in \mathcal{S} \subset \mathbb{R}^p$ ,  $p \geq 1$ . This is implemented through modeling the conditional Fréchet mean  $\mathbb{E}_{\oplus}[\{\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(t) : t \in [0, 1]\} | Z_i = z]$  through applying a global Fréchet regression step (Petersen and Müller, 2019).

Since a geodesic is determined by the two endpoints, the geodesic path  $\{\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(t) : t \in [0, 1]\}$  can be represented as a  $\mathcal{M}$ -valued pair  $(\nu_{i0}, \nu_{i1}) \in (\mathcal{D}_{\mathcal{M}}, d_{\mathcal{M}})$ , where the space  $(\mathcal{D}_{\mathcal{M}}, d_{\mathcal{M}})$  is the product metric space  $(\mathcal{M}, d) \times (\mathcal{M}, d)$  with the metric

$$d_{\mathcal{M}}((a_1, b_1), (a_2, b_2)) := \sqrt{d^2(a_1, a_2) + d^2(b_1, b_2)}, \quad \text{for all } a_1, a_2, b_1, b_2 \in (\mathcal{M}, d). \quad (6)$$

In the context of metric geometry such product metric spaces with a  $l_2$ -type metric that combines the metrics of the original spaces have been extensively studied. In particular, it is well known that  $\mathcal{D}_{\mathcal{M}}$  is a geodesic space if and only if  $\mathcal{M}$  is geodesic (Burago et al., 2001). This decomposition enables us to model the effective object response pair separately as

$$\begin{aligned} \zeta_{\oplus}(z) &= \mathbb{E}_{\oplus}[(\nu_{i0}, \nu_{i1}) | Z_i = z] = \underset{(\mu_1, \mu_2) \in (\mathcal{D}_{\mathcal{M}}, d_{\mathcal{M}})}{\operatorname{argmin}} \mathbb{E}[d_{\mathcal{M}}^2((\mu_1, \mu_2), (\nu_{i0}, \nu_{i1})) | Z_i = z] \\ &= \underset{(\mu_1, \mu_2) \in (\mathcal{D}_{\mathcal{M}}, d_{\mathcal{M}})}{\operatorname{argmin}} \mathbb{E}[d^2(\mu_1, \nu_{i0}) + d^2(\mu_2, \nu_{i1}) | Z_i = z]. \end{aligned} \quad (7)$$

This optimization problem is separable with optimal solution  $\zeta_{\oplus}(z) = (\zeta_0(z), \zeta_1(z))^\top$  where

$$\zeta_0(z) = \operatorname{argmin}_{\mu_1 \in \mathcal{M}} \mathbb{E} [d^2(\mu_1, \nu_{i0}) | Z_i = z], \quad \zeta_1(z) = \operatorname{argmin}_{\mu_2 \in \mathcal{M}} \mathbb{E} [d^2(\mu_2, \nu_{i1}) | Z_i = z]. \quad (8)$$

To implement the second step regression for higher dimensional predictors  $Z \in \mathcal{S} \subset \mathbb{R}^p$ ,  $p \geq 2$ , we use the global Fréchet regression (GFR) (Petersen and Müller, 2019) method, which is a generalization of multiple linear regression for random object responses, and thus provides a direct extension of the multiple linear regression step for the baseline covariate effect that is implemented in classical random effects modeling for Euclidean responses. For Euclidean data, the GFR approach is equivalent to fitting a multiple linear regression model by least squares.

Employing the GFR approach, defining a weight function  $s(Z, z) = 1 + (Z - \mu_Z)^\top \Sigma_Z^{-1} (z - \mu_Z)$  with  $\mu_Z = \mathbb{E}(Z)$  and  $\Sigma_Z = \operatorname{var}(Z)$ , the regression step in model (8) can be written as  $\zeta_{\oplus}(z) = (\zeta_0(z), \zeta_1(z))^\top$ , where

$$\zeta_k(z) = \operatorname{argmin}_{\mu \in \mathcal{M}} \mathbb{E} [s(Z, z) d^2(\mu, \nu_{ik})], \quad k = 0, 1. \quad (9)$$

Combining a subject-specific approach in model (5) with model (8) for the impact of the covariate  $Z$  thus provides a direct generalization of the standard random effects model (1).

### 3 Estimation and theory

Consider the global Fréchet regression (GFR) model with a response  $Y \in (\mathcal{M}, d)$  and a predictor  $T \in \mathcal{T} \subset [0, 1]$  given by

$$m_{\oplus}(t) = \operatorname{argmin}_{\mu \in \mathcal{M}} \mathbb{E} [w(T, t) d^2(\mu, Y)], \quad (10)$$

where  $w(T, t) = 1 + (T - \mu_T)^\top \Sigma_T^{-1} (t - \mu_T)$  are weight functions that are linear in  $t$ , with  $\mu_T = \mathbb{E}(T)$  and  $\Sigma_T = \operatorname{var}(T)$ . Based on the observations  $(Y_{ij}, T_{ij})$   $j = 1, \dots, n_i$ , for any given subject  $i$ ,  $i = 1, \dots, n$ ,

following (10), a subject-specific version of the GFR model is

$$m_{i\oplus}(t) = \operatorname{argmin}_{\mu \in \mathcal{M}} \mathbb{E} [w(T_{ij}, t) d^2(\mu, Y_{ij})], \quad (11)$$

where the weight function  $w$  is defined as before. This model will be implemented to recover individual trajectories from the data available separately for each subject, where we first assume the data lie exactly on the underlying geodesic and subsequently consider the small error case, dealing with additional perturbations of the responses.

Using a similar idea as Theorem 1 of Fan and Müller (2021) the following result shows that in the noise-free case the geodesic paths coincide with the GFR path.

**Theorem 1.** *Consider the sample  $(T_{ij}, Y_{ij})$ ,  $T_{ij} \in [0, 1]$   $j = 1, \dots, n_i$ . For each subject  $i$  assume that there exists a geodesic  $\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(t) \in (\mathcal{M}, d)$ ,  $t \in [0, 1]$  that uniquely connects the endpoints  $\nu_{i0} = \gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(0)$  and  $\nu_{i1} = \gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(1)$  such that the responses  $Y_{ij} = Y_i(T_{ij})$  are located exactly on this geodesic, that is, for each  $Y_{ij} \in (\mathcal{M}, d)$  there exists a  $u_{ij} \in (0, 1)$  with  $Y_{ij} = \gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(u_{ij})$ . If the predictors  $T_{ij}$  for any given subject  $i$  are such that  $T_{ij} = au_{ij} + b$ ,  $j = 1, \dots, n_i$ , for some constants  $a, b \in \mathbb{R}$ , implementing the global Fréchet regression in (11) exactly recovers the geodesic  $\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(t) : t \in [0, 1]$ . If the geodesic is extendable from  $[0, 1]$  to  $[s_1, s_2]$  and the extension is unique in the sense that it is the only geodesic connecting  $\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(s_1)$  and  $\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(s_2)$ , then the global Fréchet regression recovers the extended geodesic.*

Under the assumptions of Theorem 1 the GFR path  $\{m_{i\oplus}(t) : t \in [0, 1]\}$  coincides with the underlying geodesic path  $\{\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(t) : t \in [0, 1]\}$  and the latter can be represented by the two endpoints  $(\nu_{i0}, \nu_{i1})$  with  $m_{i\oplus}(t) = \gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(t)$  for  $t = 0, 1$ . If the geodesics are uniquely extendable, the pair  $(m_{i\oplus}(0), m_{i\oplus}(1))$  effectively represents the  $\mathcal{M}$ -valued random effect for the  $i$ -th subject and therefore serves as response for a second Fréchet regression as per model (7), (8), with the covariate  $Z$  as predictor.

In practical implementation, we replace  $(m_{i\oplus}(0), m_{i\oplus}(1))$  by the empirical version of GFR

$$\hat{m}_{i\oplus}(t) = \operatorname{argmin}_{\mu \in \mathcal{M}} \frac{1}{n_i} \sum_{j=1}^{n_i} w(T_{ij}, t) d^2(\mu, Y_{ij}), \quad t = 0, 1, \quad (12)$$

where the empirical weights are  $w = 1 + (T_{ij} - \bar{T}_i)^\top \hat{\Sigma}_{T_i}^{-1}(t - \bar{T}_i)$ , with  $\bar{T}_i$  and  $\hat{\Sigma}_{T_i}$  being the sample

mean and covariance matrix for the predictor  $T_{ij}$ ,  $j = 1, \dots, n_i$  for the  $i^{\text{th}}$  subject. With estimated object responses  $(\hat{m}_{i\oplus}(0), \hat{m}_{i\oplus}(1))$  in hand, we proceed with the GFR implementation to recover the effect of covariates  $Z$ , where  $\hat{\zeta}_{\oplus}(z) = (\hat{\zeta}_0(z), \hat{\zeta}_1(z))^{\top}$  and

$$\hat{\zeta}_k(z) = \underset{\mu_1 \in \mathcal{M}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n s_{in}(Z_i, z) d^2(\mu_1, \hat{m}_{i\oplus}(k)), \quad k = 0, 1, \quad (13)$$

where the empirical GFR weights are given by

$$s_{in}(Z_i, z) = 1 + (Z_i - \bar{Z})^{\top} \hat{\Sigma}_Z^{-1} (z - \bar{Z}), \quad (14)$$

$\bar{Z}$  and  $\hat{\Sigma}_Z$  being the sample mean and covariance matrix for the predictor  $Z_i$ ,  $i = 1, \dots, n$ .

**Theorem 2.** *Under assumptions (R0)-(R2) in the Appendix it holds that*

$$d_{\mathcal{M}}(\hat{\zeta}_{\oplus}(z), \zeta_{\oplus}(z)) = O_P(n^{-1/2}).$$

Next we discuss the more realistic case where responses do not lie exactly on the underlying geodesic paths but instead are perturbed from those on the path as per (2), (4). To this end, let  $(\Omega^*, \mathcal{F}^*, P^*)$  be the underlying probability space on which the observed data  $(T_{ij}, Y_{ij})$  are defined for the  $i^{\text{th}}$  subject,  $i = 1, \dots, n$ ,  $j = 1, \dots, n_i$ . Since the mechanism that generates the data are independent as per (A1),  $(\Omega^*, \mathcal{F}^*, P^*)$  can be perceived as a product space of two probability spaces:  $(\Omega_1, \mathcal{F}_1, P_1)$ , on which the  $\mathcal{M}$ -valued geodesic  $\{\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(t) : t \in [0, 1]\}$  connecting the two points  $\nu_{i0}$  and  $\nu_{i1}$ , is defined; and  $(\Omega_2, \mathcal{F}_2, P_2)$ , on which the observed time points  $T_{ij}$  and the random perturbation maps  $\mathcal{P}_{ij}$  associated with the noisy observation  $Y_{ij}$  are defined. Thus, one can attribute the randomness of the noisy observations to three sources,  $Y_{ij} = \mathcal{P}_{ij} \left[ \gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(T_{ij}) \right] = f(\omega_1, \omega_2, \omega_3)$ , where  $\omega_1$  is a random element in  $(\Omega_1, \mathcal{F}_1, P_1)$  that generates the endpoints of the true geodesic trajectory, thus generating the underlying geodesic;  $(\omega_2, \omega_3) \in (\Omega_2, \mathcal{F}_2, P_2)$ , where  $\omega_2$  generates the  $T_{ij}$  and  $\omega_3$  generates  $\mathcal{P}_{ij}$  for  $j = 1, \dots, n_i$ ;  $i = 1 \dots, n$ . For the special case of random effects models in Euclidean space,  $(\Omega_1, \mathcal{F}_1, P_1)$  is the underlying probability space for random slope and intercept.

Note that fixing some element  $\omega_1 \in \Omega_1$  corresponds to a realization of the  $\mathcal{M}$ -valued underlying geodesic process. Also, as per assumption (A1), given a  $\omega_1 \in (\Omega_1, \mathcal{F}_1, P_1)$ ,  $(T_{\cdot j}, \mathcal{P}_{\cdot j})$  are independent

in  $(\Omega_2, \mathcal{F}_2, P_2)$  for all  $j = 1, \dots, n_i$  and do not depend on  $\omega_1$ . Suppose that for a given  $\omega_1 \in (\Omega_1, \mathcal{F}_1, P_1)$ , the geodesic  $\gamma_{\nu_0, \nu_1}^{(\omega_1)}$  is observed at  $m$  random time points.

We use notations  $\gamma_{\nu_0, \nu_1}^{(\omega_1)}(\cdot)$ ,  $\mathcal{P}(\gamma_{\nu_0, \nu_1}^{(\omega_1)}(\cdot))$  and  $T$  to represent the corresponding quantities for the underlying geodesic, noisy observation and the random time point, respectively, for any given  $\omega_1 \in (\Omega_1, \mathcal{F}_1, P_1)$ . Denote by  $\mathbb{E}_{\Omega_2}$  the expectation with respect to the probability measure  $P_2$ . For any  $t \in [0, 1]$ , define  $\gamma_{\nu_0, \nu_1}^{(\omega_1)}(t) = \operatorname{argmin}_{\mu \in \mathcal{M}} \mathbb{E}_{\Omega_2} \left[ d^2(\mathcal{P}(\gamma_{\nu_0, \nu_1}^{(\omega_1)}(\cdot)), \mu) | T = t \right]$ . We make the following small errors assumption, which mean that errors implemented in the form of perturbations are asymptotically negligible, uniformly across all realizations of the geodesic paths,

$$(A2) \quad \mathbb{E}_{\Omega_2} \left[ d^2 \left( \mathcal{P} \left( \gamma_{\nu_0, \nu_1}^{(\omega_1)}(T) \right), \gamma_{\nu_0, \nu_1}^{(\omega_1)}(T) \right) \right] = O(\alpha_n^2), \text{ with } \alpha_n \rightarrow 0 \text{ and } n\alpha_n^2 \rightarrow \infty.$$

For classical Euclidean linear random effects models with an additive error structure, this small errors assumption is not required due to the availability of additive operations, permitting the application of the law of large numbers and central limit theorem. None of these is available in general geodesic spaces. A small error assumption is commonly required in nonlinear models with measurement errors and instrumental variable models (Amemiya, 1985; Chen et al., 2011; Carroll and Hall, 2004; Carroll et al., 2006; Schennach, 2016). Observing that the underlying true geodesic trajectory for the  $i^{\text{th}}$  individual given by  $\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(\cdot)$  is a random realization corresponding to some  $\omega_1$  in the probability space  $(\Omega_1, \mathcal{F}_1, P_1)$ , define the GFR model at the population level for any  $\omega_1 \in (\Omega_1, \mathcal{F}_1, P_1)$  as

$$\mu_{\omega_1}^*(t) = \operatorname{argmin}_{\mu \in \mathcal{M}} H_{\omega_1}(\mu, t), \quad H_{\omega_1}(\mu, t) = \mathbb{E}_{\Omega_2} \left[ w(T, t) d^2(\gamma_{\nu_0, \nu_1}^{(\omega_1)}(T), \mu) \right], \quad (15)$$

where  $w(T, t) = 1 + (T - \mu_T)^\top \Sigma_T^{-1} (t - \mu_T)$  is the GFR weight function with  $\mu_T = \mathbb{E}(T)$  and  $\Sigma_T = \operatorname{var}(T)$ , as before, and  $H_{\omega_1}$  would be the objective function to minimize using global Fréchet regression with a fixed target response on the geodesic for a given  $\omega_1 \in (\Omega_1, \mathcal{F}_1, P_1)$ , if there was no error in the observations. Since in the error-free case the GFR path recovers the geodesic entirely,  $\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(t)$  equals  $\mu_{\omega_1}^*(t)$ ,  $t \in \mathcal{T}$  for some  $\omega_1 \in (\Omega_1, \mathcal{F}_1, P_1)$ ,  $i = 1, \dots, n$ .

On the other hand, a GFR model based on the observed noisy responses, for any given  $\omega_1 \in$

$(\Omega_1, \mathcal{F}_1, P_1)$ , can be defined as

$$\tilde{\mu}_{\omega_1}(t) = \operatorname{argmin}_{\mu \in \mathcal{M}} G_{\omega_1}(\mu, t), \quad G_{\omega_1}(\mu, t) = \mathbb{E}_{\Omega_2} [w(T, t) d^2(\mathcal{P}(\gamma_{\nu_0, \nu_1}^{(\omega_1)}(\cdot))(T), \mu)], \quad (16)$$

where the weight function  $w(T, t)$  for the global Fréchet regression is defined as before. In our notation the GFR path  $m_{i\oplus}(t) \in \mathcal{M}$  for the  $i^{\text{th}}$  subject corresponds to  $\tilde{\mu}_{\omega_1}(t), t \in \mathcal{T}$  for some  $\omega_1 \in (\Omega_1, \mathcal{F}_1, P_1)$ . In other words, the quantities  $\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(\cdot)$  and  $m_{i\oplus}(\cdot)$ , for  $i = 1, \dots, n$ , are the subject-level realizations of  $\mu_{\omega_1}^*(\cdot)$  and  $\tilde{\mu}_{\omega_1}(\cdot)$ , respectively for some random element  $\omega_1 \in (\Omega_1, \mathcal{F}_1, P_1)$ . We require the following assumptions for all  $\omega_1 \in (\Omega_1, \mathcal{F}_1, P_1)$ .

(K1) For any given  $t \in \mathcal{T}$ , the Fréchet means  $\mu_{\omega_1}^*(t)$  and  $\tilde{\mu}_{\omega_1}(t)$  exist and are unique, and for any

$$\varepsilon > 0 \text{ it holds that } \inf_{d(\mu_{\omega_1}^*(t), \mu) > \varepsilon} [H_{\omega_1}(\mu, t) - H_{\omega_1}(\mu_{\omega_1}^*(t), t)] > 0$$

and  $\inf_{d(\tilde{\mu}_{\omega_1}(t), \mu) > \varepsilon} [G_{\omega_1}(\mu, t) - G_{\omega_1}(\tilde{\mu}_{\omega_1}(t), t)] > 0.$

(K2) There exist constants  $C_1 > 0, \beta_1 > 1$ , such that for all  $\eta > 0$ ,

$$\inf_{\omega_1 \in (\Omega_1, \mathcal{F}_1, P_1)} \inf_{d(\mu_{\omega_1}^*(t), \mu) < \eta} [H_{\omega_1}(\mu, t) - H_{\omega_1}(\mu_{\omega_1}^*(t), t) - C_1 d(\mu_{\omega_1}^*(t), \mu)^{\beta_1}] \geq 0.$$

Assumption (K1) is commonly used to establish consistency of an M -estimator (see Chapter 3.2 in [Van der Vaart and Wellner \(2000\)](#)). It ensures weak convergence of the empirical process  $H_{\omega_1} - G_{\omega_1}$ , which in turn implies convergence of the minimizers ([Chen and Müller, 2022](#)). Assumption (K2) relates to the curvature of the objective function and is needed to control the behavior of the true and perturbed objective functions  $H_{\omega_1}$  and  $G_{\omega_1}$ , respectively, near their minimizers. These assumptions are satisfied for many random objects of interest ([Petersen and Müller, 2019](#)).

The following lemma establishes a connection between the population level Fréchet means of the responses lying exactly on a geodesic (given in (15)) and the perturbed responses situated near but not on the geodesic (given in (16)).

**Lemma 1.** *Under assumptions (A1), (A2) and (K1), (K2), for any given  $t \in \mathcal{T}$ ,*

$$\sup_{\omega_1 \in (\Omega_1, \mathcal{F}_1, P_1)} d(\mu_{\omega_1}^*(t), \tilde{\mu}_{\omega_1}(t)) = O(\alpha_n),$$

where  $\alpha_n$  is as defined in Assumption (A2). Further, for any  $i = 1, \dots, n$  and any  $t \in \mathcal{T}$ ,

$$d(m_{i\oplus}(t), \gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(t)) = O(\alpha_n).$$

The above lemma implies that for any individual  $i$ , the underlying geodesic trajectory  $\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(\cdot)$  can be recovered pointwise with asymptotically negligible error by the GFR path for the  $i^{\text{th}}$  individual arbitrarily closely;  $i = 1, \dots, n$ . This suggests to estimate the underlying subject-specific geodesic from the noisy observations for each subject by the same method as before, obtaining the GFR path as in (12). Pointwise consistency of estimates is sufficient as one only needs to recover the endpoints  $(\nu_{i0}, \nu_{i1})$  of the geodesic. We follow the same approach as before to infer the effect of the covariate  $Z$  by implementing (13). This is justified by the following result, which provides the rate of convergence of the regression of the metric space-valued random effects on the covariate  $Z$ .

**Theorem 3.** *Under assumptions (A1), (A2), (K1), (K2) and (R0), (R2) in the Appendix, for any  $z \in \mathcal{S} \subset \mathbb{R}^p$  with  $p \geq 1$ ,*

$$d_{\mathcal{M}}\left(\hat{\zeta}_{\oplus}(z), \zeta_{\oplus}(z)\right) = O_P(\alpha_n^{1/2}).$$

From the definition of  $\alpha_n$  in assumption (A2), the rate is slightly slower than  $n^{-1/4}$ .

## 4 Simulation studies

We report here only a subset of our simulation results for the important case of responses in the space of univariate distributions endowed with the Wasserstein metric, while additional simulation results for spherical data can be found in Section 6 in the Supplement.

The Wasserstein space of probability distributions that we consider here is as in Example 1, with time-varying distributions as responses  $Y$  which can be represented by their quantile functions  $Q_Y(\cdot)$ . For each subject  $i$ , the random responses are constructed as repeated measurements around some underlying geodesic path  $\gamma_{\nu_{i0}, \nu_{i1}}^{(i)}(\cdot)$  in the space  $(\mathcal{W}_2(M), d_W)$ . These underlying geodesic paths were generated conditional on a covariate  $Z_i \in \mathcal{S} \subset \mathbb{R}$ , while the observed responses were sampled on these geodesics and then perturbed, implementing the following steps. For each subject

**Step 1.** Generate  $Z_i \stackrel{i.i.d.}{\sim} \text{Unif}(-1, 1)$ .

**Step 2.** Generate  $n_i$  random time points  $T_{ij} \stackrel{i.i.d.}{\sim} \text{Unif}(0, 1)$ . We consider a dense design, where  $n_i = 50$ , as well as a sparse design, where  $n_i \in \{2, \dots, 5\}$  with equal probability.

**Step 3.** Generate end points of the geodesics,  $\nu_{i0}$  and  $\nu_{i1}$ , that depend on the external covariate  $Z_i$  in the following way. Representing  $\nu_{i0}$  and  $\nu_{i1}$  as quantile functions  $Q_{\nu_{i0}}(\cdot)$  and  $Q_{\nu_{i1}}(\cdot)$ , the conditional expectation of  $\nu_{ik}$  given  $Z$  is modeled as

$$\mathbb{E}[Q_{\nu_{ik}}(\cdot)|Z_i = z, T_{ij} = u] = \xi_{u,z} + \sigma_{u,z} \Phi_{[0,1]}^{-1}(\cdot), \quad k = 0, 1, \quad j = 1, \dots, n_i, \quad (17)$$

where  $\Phi_{[0,1]}(\cdot)$  is the cdf of a standard normal distribution truncated on  $[0, 1]$ . Specifically, the corresponding distribution function is given by

$$F(x) = \frac{\Phi((x - \xi_{u,z})/\sigma_{u,z}) - \Phi(-\xi_{u,z}/\sigma_{u,z})}{\Phi((1 - \xi_{u,z})/\sigma_{u,z}) - \Phi(-\xi_{u,z}/\sigma_{u,z})} \mathbf{1}_{[0,1]}(x) + \mathbf{1}_{(1,\infty)}(x), \quad x \in \mathbb{R}.$$

The distributional responses  $\nu_{i0}$  and  $\nu_{i1}$  are perturbed versions from model (17). We consider four different simulation scenarios for location-scale families with varying sample sizes and perturbation levels, for both sparse and dense sample designs. The global parameters considered in the following data generation mechanisms are  $\mu_0 = 0$ ,  $\sigma_0 = 0.1$ ,  $\beta_1 = 0.3$ ,  $\beta_2 = 0.25$ ,  $\gamma = 0.3$ ,  $\nu_1 = 0.25$ ,  $\nu_2 = 1$ .

**Setting I.** The mean changes with the predictor values while the variance is constant. We generate the the auxiliary distribution parameters independently as  $\mu_Y|(Z = z, T = u) \sim N_{[0,1]}(\xi_{u,z}, \nu_1)$  and  $\sigma_Y|(Z = z, T = u) = \sigma_{u,z}$ , where  $\xi_{u,z} = \mu_0 + \beta_1 z + \beta_2 u$  and  $\sigma_{u,z} = 0.1$ . The corresponding distribution is given by  $Q_{\nu_{ik}} = \mu_Y + \sigma_Y \Phi^{-1}$ ,  $k = 0, 1$ , where  $\Phi$  is the standard normal cdf.

**Setting II.** The mean remains constant, while the variance changes w.r.t to the predictor values. Again, the distribution parameters are generated independently as  $\mu_Y|(Z = z, T = u) \sim N_{[0,1]}(\xi_{u,z}, \nu_1)$  and  $\sigma_Y|(Z = z, T = u) \sim \text{Gamma}(\frac{(\sigma_0 + \gamma z)^2}{\nu_2}, \frac{\nu_2}{(\sigma_0 + \gamma z)^2})$ , such that  $\xi_{u,z} = \mu_0 + \beta_2 u$  and  $\sigma_{u,z} = \sigma_0 + \gamma z$ .

**Setting III.** The mean and variance both vary w.r.t to the predictor values. To this end,  $\mu_Y|(Z = z, T = u) \sim N_{[0,1]}(\xi_{u,z}, \nu_1)$  and  $\sigma_Y|(Z = z, T = u) \sim \text{Gamma}(\frac{(\sigma_0 + \gamma z)^2}{\nu_2}, \frac{\nu_2}{(\sigma_0 + \gamma z)^2})$ , independently sampled such that  $\xi_{u,z} = \mu_0 + \beta_1 z + \beta_2 u$  and  $\sigma_{u,z} = \sigma_0 + \gamma z$ .

**Setting IV.** After sampling the distribution parameters as in the previous setting, the resulting distribution is then “transported” in Wasserstein space via a random transport map  $T$ , that is

uniformly sampled from a family of perturbation/ distortion functions  $\{T_k : k \in \pm 1, \pm 2, \pm 3\}$ , where  $T_k(a) = a - \frac{\sin(\pi ka)}{|k\pi|}$ . The transported distribution is given by  $T\#(\xi_{u,z} + \sigma_{u,z}\Phi_{[0,1]}^{-1}(\cdot))$ , where  $T\#p$  is a push-forward measure such that  $T\#p(A) = p(\{x : T(x) \in A\})$ , for any measurable function  $T : \mathbb{R} \rightarrow \mathbb{R}$ , distribution  $p \in \mathcal{W}$ , and set  $A \subset \mathbb{R}$ . We sample the random transport map  $T$  uniformly from the collection of maps described above;  $p$  denotes a truncated Gaussian distribution with parameters  $\xi_{u,z}$  and  $\sigma_{u,z}$ , and  $\mathcal{W}$  is the metric space of distributions equipped with the Wasserstein metric. The distributions thus generated are not Gaussian anymore due to the transportation. The Fréchet mean can be shown to remain at  $\xi_{u,z} + \sigma_{u,z}\Phi^{-1}(\cdot)$  as before. Then the geodesic in the quantile space connecting  $Q_{\nu_{i0}}(\cdot)$  and  $Q_{\nu_{i1}}(\cdot)$  is given by  $Q_{\gamma_{\nu_{i0},\nu_{i1}}^{(i)}(\cdot)} : t \mapsto (1-t)Q_{\nu_{i0}} + tQ_{\nu_{i1}}$ . For the  $i^{\text{th}}$  subject,  $n_i$  points are generated on the true underlying geodesic  $Q_{\gamma_{\nu_{i0},\nu_{i1}}^{(i)}(t)}$ ,  $t \in [0, 1]$ .

**Step 4.** Perturb the true quantile functions  $Q_{\gamma_{\nu_{i0},\nu_{i1}}^{(i)}(T_{ij})}$  situated on a geodesic such that the observed responses remain valid quantile functions. The perturbed/noisy distributional responses, represented as quantile functions, are constructed as  $\tilde{Q} : [0, 1] \rightarrow [0, 1]$  such that

$$\tilde{Q}(u) = Q(s) + \varepsilon\Delta(s), \quad s \in [0, 1], \quad (18)$$

where  $\Delta(s) = \alpha_n Q(s)(1 - Q(s))$ ,  $0 < \alpha_n < 1$  and  $\varepsilon = \pm\alpha_n$  with equal probability 1/2. With a sufficiently small choice of  $\alpha_n \in (0, 1)$ ,  $\tilde{Q}$  is an increasing quantile function in  $[0, 1]$ . From the construction we have  $\Delta(s) \leq \min\{Q(s), 1 - Q(s)\}$  for all  $s \in [0, 1]$ . Note that for  $0 < \alpha_n < 1$ ,  $\tilde{Q}' = (Q \pm \alpha_n \Delta)' > 0$ , as long as the true quantile functions  $Q$  are strictly increasing and  $\mathbb{E}(\tilde{Q}) = Q$ . The observed responses are thus per (18)  $\tilde{Q}_{Y_{ij}(T_{ij})} = Q_{\gamma_{\nu_{i0},\nu_{i1}}^{(i)}(T_{ij})} \pm \alpha_n^2 Q_{\gamma_{\nu_{i0},\nu_{i1}}^{(i)}(T_{ij})} (1 - Q_{\gamma_{\nu_{i0},\nu_{i1}}^{(i)}(T_{ij})})$ . We implemented the proposed model as per (12) and (13).

The effect of the perturbation parameter  $\alpha_n$  is demonstrated in Figure 1 for one simulation run in Setting IV. True, observed, and predicted distributions are shown for the sparse design case. The predicted distributions are obtained for the observed values of the covariate/predictor  $Z_i$  for all  $t \in [0, 1]$ , represented as densities. For small perturbations, the observed distributions are seen to be close to the underlying true geodesic path of distributions, while for larger levels of perturbation deviations are larger. However, estimated/predicted distributions throughout remain close to the true distributions, providing evidence for the efficiency of the proposed random effects model.

We illustrate the effects of the covariate  $Z$  on the model fits across different simulation settings

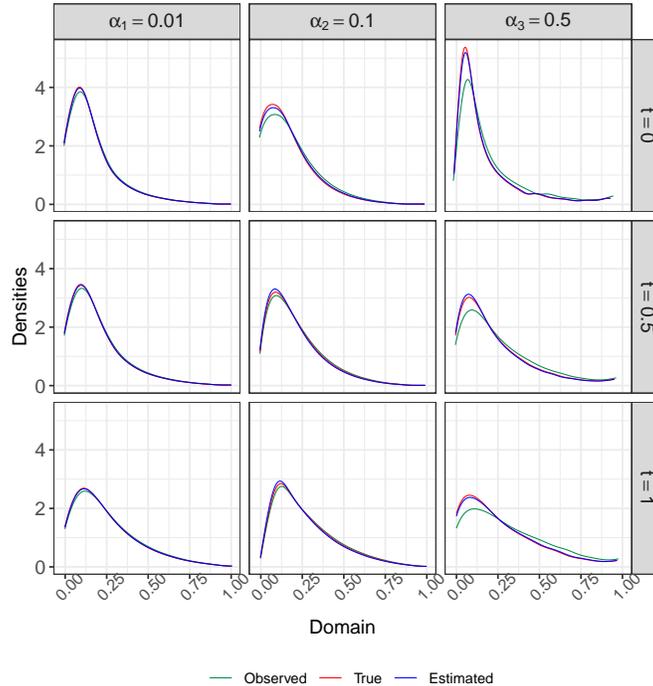


Figure 1: Visualization of the true (red), observed (perturbed, green), and estimated (blue) distributional object responses as densities for a randomly selected simulation sample generated under setting IV with a sparse design where each subject has 2 to 5 repeated measurements, comparing varying perturbation levels  $\alpha_n = 0.01, 0.1, 0.3$  (left, middle and right). The densities lying on a geodesic in the Wasserstein space of distributions are displayed at three different time points,  $t = 0, 0.5$ , and 1 (top, middle, and bottom rows, respectively).

for one simulation run in Figure 2. Again data are generated for a sparse design for each of the settings mentioned above with sample size  $n = 500$ , where the observed distributions are generated around the true underlying geodesics in the Wasserstein space and observations are perturbed at perturbation level  $\alpha_n = 0.1$ . To assess the covariate effects, we fitted the model at covariate levels that correspond to the 10%, 50%, and 90% quantiles of the covariate  $Z$ . One finds that at all time points and across all settings the predicted densities closely approximate the truth. This demonstrates that in the small error case the proposed random effects model and its implementation is well suited to recover the true trajectories when given the covariate information. For setting I with a location shift in the data generation mechanism, the modes of the densities shift towards the right, i.e., a higher value of the covariate is associated with a right shift in the estimated densities. For setting II, a higher value of the covariate is associated with an increase in the spread of the distribution. Settings III and IV capture the combined effect of location and scale shifts. The location, spread, skewness, and overall shape of the predicted densities change as expected with

increasing levels of the predictor values. One also observes a rightwards shift of the distributions over time, an expected consequence of the generation of the geodesics in distribution space.

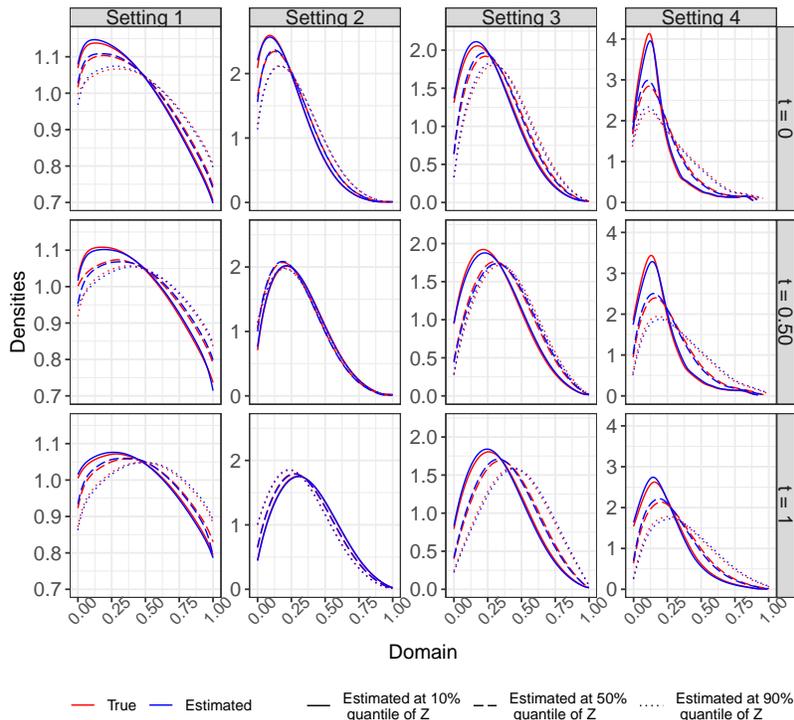


Figure 2: The time-dynamic effect of the baseline covariate for distributional objects represented as densities for a randomly selected simulation sample, displaying true (red) and estimated (blue) densities for simulation settings I-IV (columns from left to right). Data were generated under a sparse design, where each subject has 2 to 5 repeated measurements and where response distributions were perturbed with a fixed small perturbation level  $\alpha = 0.1$ . Estimated/predicted densities are shown for the 10% (solid), 50% (long-dashed) and 90% (dotted) quantile levels of the covariate. The top, middle, and bottom panels correspond to the prediction/estimation at times  $t = 0, 0.5$ , and 1, respectively.

We further studied the effect of sample size and sample design (sparse or dense) for the four simulation settings on the performance of the proposed method while keeping the perturbation level fixed at  $\alpha = 0.1$ . The results of 500 Monte Carlo simulation runs are shown in Figure 3, where we display boxplots of Integrated Square Error (ISE) as a measure of discrepancy between the true and the estimated distributions. Specifically,

$$\text{ISE}_r = \int_{z \in \mathcal{S}} \int_{t \in [0,1]} d_W(Y^r(t, z), \hat{Y}^r(t, z)) dt dz, \quad (19)$$

where  $Y^r(t, z)$  and  $\hat{Y}^r(t, z)$  denote, respectively, the true distributional object lying on a geodesic

(without perturbation) in the Wasserstein-2 space and the estimated object at time point  $t$  and covariate value  $z$  for the  $r^{\text{th}}$  simulation run, where  $r = 1, \dots, 500$ . We observe a decrease in ISE for

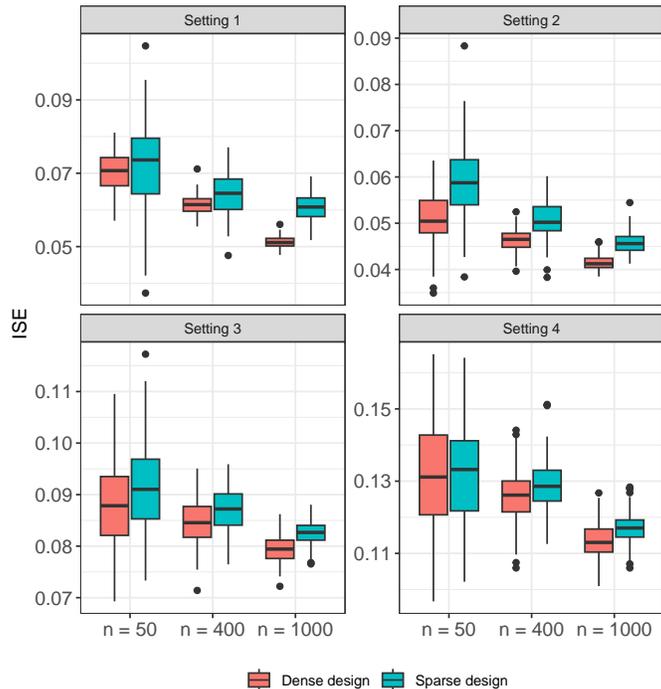


Figure 3: Boxplots of Integrated Squared Errors (ISE) calculated as per (19), over 500 simulation runs for the four simulation settings (displayed in the panels clockwise from the top left corner). Results are shown for sample sizes  $n = 50, 400, 1000$  for both sparse (blue) and dense (red) designs.

increasing sample size and deviations are generally higher if both the location and scale parameters are varied as a function of the covariate.

#### 4.1 Simulation study: Responses lying on the surface of a sphere

We applied the proposed approach targeting general random objects as responses lying on the surface of a sphere. The numerical results describing the data generation mechanism and evaluating the performance of the proposed method are discussed in details in subsection 4.1 of the Supplement.

## 5 Data analysis

### 5.1 Longitudinal fMRI data

Resting-state functional Magnetic Resonance Imaging (fMRI) methodology makes it possible to study brain activation and to identify brain regions or cortical hubs that exhibit similar activity when subjects are in the resting state (Allen et al., 2014). fMRI measures brain activity by detecting changes in blood-oxygen-level-dependent (BOLD) signals in the brain across time. The analysis of brain functional connectivity at the subject level typically relies on a specific spatial parcellation of the brain into a set of regions of interest (ROIs). Temporal coherence between pairwise ROIs is usually measured by the so-called Pearson correlation coefficient matrix (PCC) of functional connectivity obtained from the fMRI time series, which is an  $m \times m$  correlation matrix if one has  $m$  distinct ROIs. In this analysis, we will use PCC matrices derived from fMRI as responses. Alzheimer’s Disease has been found to be associated with anomalies in the functional integration of ROIs (Damoiseaux et al., 2012; Zhang et al., 2010) that may be time-varying, along with changes in the brain due to aging for cognitively normal subjects. This provides the motivation to explore the time-varying regression relationship between the connectivity correlation matrix objects and relevant external covariates.

Available data are from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu), where PCC matrices derived from fMRI signals are observed sparsely over time for each subject in a sample of  $n = 340$  subjects composed of 155 Cognitive Normal (CN) subjects and 185 Alzheimer’s patients with mild cognitive impairment (MCI) with ages ranging from 55.7 to 94.8 years. At least 2 scans are available for each subject but not more than 9 scans, with a median of 4 scans, so these are sparsely sampled longitudinal data. We normalized the time scale of the measurements to the interval  $\mathcal{T} = [0, 1]$ , where for each subject the time at which the first scan is recorded is defined as the origin of the time scale  $t = 0$  and  $t = 1$  is 7 (9) years after the first scan for the CN (MCI) subjects. The pre-processing of the BOLD signals adopted standard procedures of slice-timing correction, head motion correction, and other standard steps. Then  $m = 90$  brain seed voxels for each subject were extracted for the ROIs of the automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) to parcellate the whole brain into 90 ROIs, with 45 ROIs in

each hemisphere, and the signals were converted to a  $90 \times 90$  PCC matrix, which corresponds to the available observation for each time point and subject.

The structure of the space of random objects always depends on the choice of the metric, which is often chosen for convenience and interpretability in the context of specific data applications. Here we endow the space of symmetric positive definite correlation matrices  $\mathcal{M}$  with the power-Euclidean metric  $d_P$  with the power  $\alpha = 1/2$  (Dryden et al., 2010),

$$d_P(A, B) = \frac{1}{\alpha} \|A^\alpha - B^\alpha\|_F \quad \text{for any } A, B \in \mathcal{M}. \quad (20)$$

Here  $S^\alpha = U\Lambda^\alpha U^\top$ , for the usual spectral decomposition of  $S = U\Lambda U^\top$  with an orthogonal matrix  $U$  and a diagonal matrix  $\Lambda$  with strictly positive entries and  $\|\cdot\|_F$  denotes the Frobenius norm. The space  $\mathcal{M}$  is a uniquely extendable geodesic space. To implement the proposed random effects model, in a first step we recovered the underlying subject-specific trajectories by estimating the matrices at the endpoints 0 and 1, and then regressed these on the covariate  $Z$ , which was chosen as the two-dimensional vector (Age, ADAS-Cog-13 score) for each subject, obtained at the time of the first scan  $t = 0$ . For Alzheimer’s studies, the ADAS-Cog-13 score (henceforth referred to as the C score) is a widely-used measure of cognitive performance. It quantifies impairments across cognitive domains (Kueper et al., 2018); higher scores indicate more serious cognitive deficiency.

To illustrate the effect of the C-score, we fix the age of the subjects at its mean level (74 years) and provide the fitted model at the 10%, 50%, and 90% quantiles of the C-score. Figure 4 demonstrates the trend for the temporal correlations for varying C-score levels at different times of the study for the CN subjects. One finds that the overall correlation strengths diminish with higher C-scores. Further, comparing the rows for each panel, we find that correlations are overall weaker at time 1 than at time 0. A similar pattern with overall weaker connections emerges for the MCI subjects (see Figure 17 in the Supplement).

To further elicit the time-varying effects of the C-score on the PCC matrix geodesics, we subtract the predicted matrices at time 0 from the predicted matrices at time 1, for each of the three covariate quantiles, separately for CN and MCI subjects. In Figure 5, the columns (from left to right) display the difference of the fitted PCC matrices at time 1 and time 0, at the  $z_1 = 10\%$ ,  $z_2 = 50\%$ , and  $z_3 = 90\%$  quantiles of the C-score, respectively, while the other covariate age is fixed at its mean

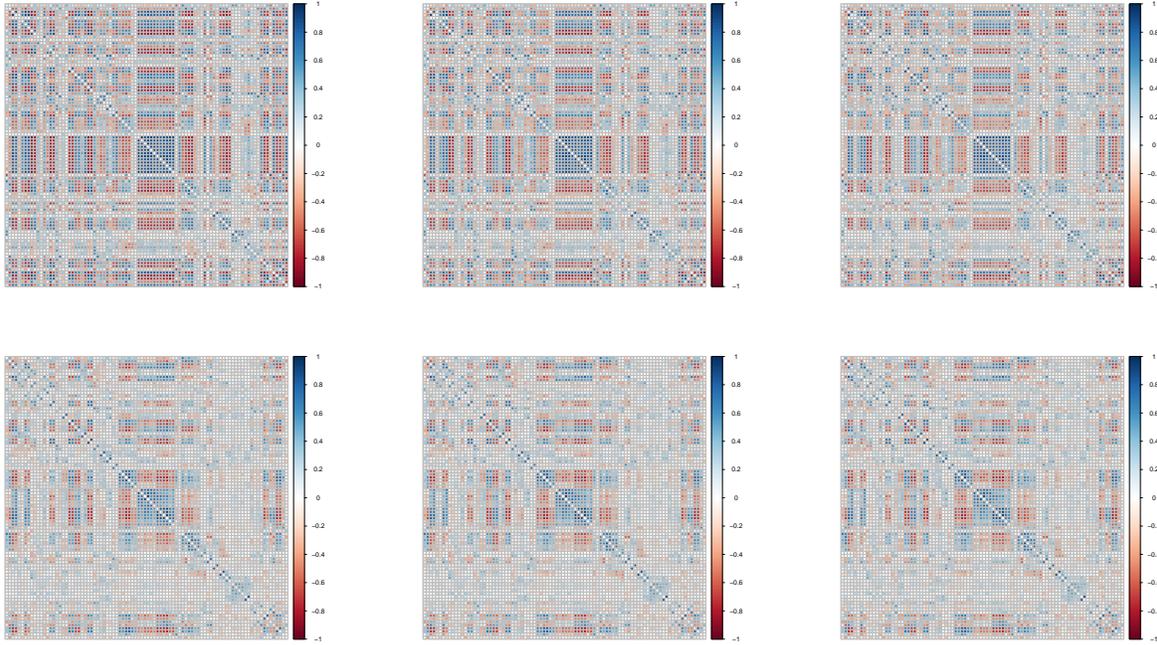


Figure 4: Estimated/predicted PCC matrices obtained from the proposed random effects model for CN subjects. The top and bottom rows correspond to the predicted correlation matrices at times  $t = 0$  and  $t = 1$  respectively, while within each row the left, middle, and right panels depict the fits at the 10%, 50%, and 90% quantiles of the C-score with age fixed at its mean level. Positive (negative) values for correlations are drawn in red (blue).

level. The top (bottom) row corresponds to the CN (MCI) subjects. For higher score levels, the inter-hub connections are found to become weaker. The effect is clearly more pronounced for the MCI subjects as compared to the CN subjects, MCI subjects losing connectivity at a faster rate.

We also converted the PCC matrices into simple, undirected, weighted networks to facilitate interpretation by setting diagonal entries to 0 and hard thresholding the absolute values of the remaining correlations. We kept the 15% strongest connections and discarded the others (Schwarz and McGonigle, 2011), converting the PCC into weighted adjacency matrices. The adjacency matrix computed from a PCC matrix is given by  $A = (a_{ij})_{i,j=1,\dots,m}$ , indicating the  $i$ -th and  $j$ -th hubs in the brain are either connected by an edge of weight  $a_{ij} > 0$ , or else unconnected if  $a_{ij} = 0$ . To represent the resulting estimated brain networks for changing covariate levels we use network summaries such as modularity, a summary measure of network segregation (Newman, 2006b) and global efficiency (Alexander-Bloch et al., 2013), a measure of network integration. With  $a_{ij}$  representing the edge weight between nodes  $i$  and  $j$ , modularity is defined as  $Q = \frac{1}{2L} \sum_{i,j} \left[ a_{ij} - \frac{k_i k_j}{2L} \right] \delta(c_i, c_j)$ , where  $L$  is the sum of all of the edge weights in the graph,  $k_i$  is the sum of the weights of the edges

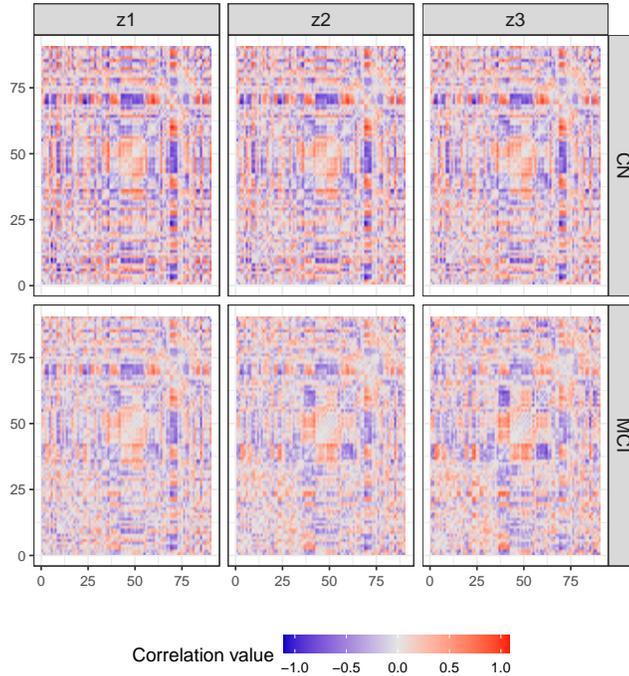


Figure 5: Differences of the predicted PCC matrices at time 1 and at time 0 for both CN (top) and MCI (bottom) subjects. The columns, from left to right, correspond to the differences of the predicted PCC matrices, fitted at the 10%, 50%, and 90% quantiles of the C-score, respectively, while the second covariate age is fixed at its mean level. Higher (lower) values of the difference are shown in red (blue).

attached to node  $i$ ,  $c_i, c_j$  are the communities of the nodes; and  $\delta(x, y) = 1$  if  $x \neq y$  and 0 otherwise. Table 1 shows modularity and global efficiency of the brain networks for CN and MCI subjects at times 0 and 1 estimated at the 10%, 50%, and 90% quantiles of the C-score, respectively, while covariate age is fixed at its mean level. Both indices decrease for higher C-scores and over time where the decrease over time is much more pronounced for MCI subjects, in line with the previous findings for PCC matrices.

We also evaluated the modularity of the predicted brain networks continuously over time between time 0 and 1 from the estimated PCC matrices on geodesics in the space of correlation matrices, see Figure 6. The modularity for the highest C-scores generally is lowest and declines throughout time, which suggests less and more rapidly declining connectivity. In contrast, modularity for low and median C-scores stays stable for a longer period, where the contrast is even higher for MCI subjects. This indicates that connectivity decline is higher for those starting with higher C-scores and lower connectivity.

The validity of the fits obtained with the proposed random effects model can be assessed by

Table 1: Modularity and global efficiency of the estimated brain networks obtained for CN and MCI subjects by hard thresholding at times  $t = 0$  and  $t = 1$ , for the 10%, 50%, and 90% quantiles of the C-score, while covariate age is fixed at its mean level.

	CN				MCI			
	Modularity		Global Efficiency		Modularity		Global Efficiency	
	$t = 0$	$t = 1$	$t = 0$	$t = 1$	$t = 0$	$t = 1$	$t = 0$	$t = 1$
10% Quantile of Total Score	0.534	0.479	0.499	0.387	0.536	0.593	0.520	0.368
50% Quantile of Total Score	0.528	0.474	0.485	0.371	0.535	0.541	0.486	0.365
90% Quantile of Total Score	0.505	0.462	0.472	0.355	0.531	0.465	0.387	0.322

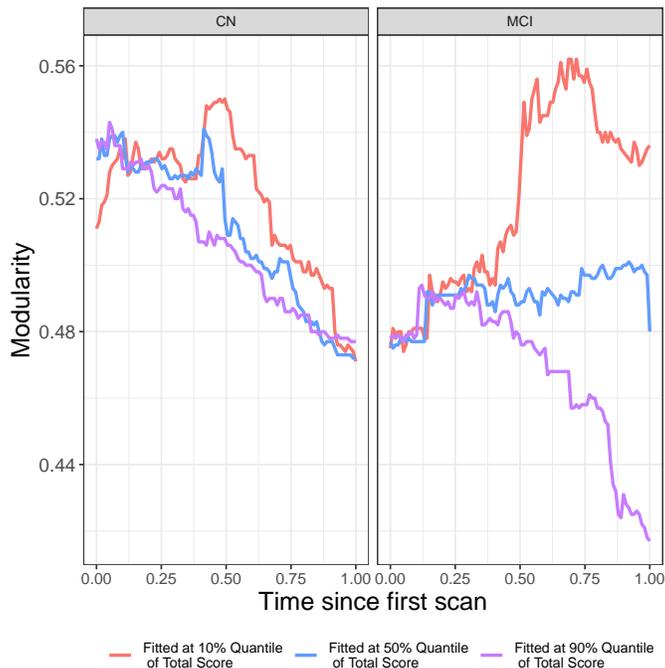


Figure 6: Modularity of the estimated brain networks over time for the CN and MCI subjects in the left and right panels, respectively). The covariate levels at which the networks are estimated are the 10% (red), 50% (blue) and 90% (purple) quantiles of the C-score, with the other covariate age fixed at its mean level.

its out-of-sample prediction performance. We randomly split the dataset into a training set with sample size  $n_{\text{train}}$  and a test set with the remaining  $n_{\text{test}}$  subjects. We then take the fitted objects  $\hat{\zeta}_{\oplus}(\cdot)$  obtained from the training set and predict the responses in the test set using the covariates present in the test set. As a measure of the efficacy of the fitted model, we compute the root mean squared prediction error

$$\text{RMPE} = \left[ \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \frac{1}{n_i} \sum_{j=1}^{n_i} d_P^2 \left( Y_{ij}^{\text{test}}, \hat{Y}_{ij}^{\text{test}} \right) \right]^{-1/2},$$

where  $Y_{ij}^{\text{test}} = Y_{ij}^{\text{test}}(T_{ij})$  denotes the  $i^{\text{th}}$  observed response at time  $T_{ij}$ ,  $j = 1, \dots, n_i$ , for the  $i^{\text{th}}$  subject in the test set,  $\hat{Y}_{ij}^{\text{test}} = \hat{Y}_{ij}^{\text{test}}(T_{ij})$  denotes the predicted object at the covariate level  $Z_i$  for the predictors in the test set, and  $d_P$  the power metric in  $\mathcal{M}$ , with power  $\alpha = 1/2$ . We repeat this process 100 times and compute the RMPE for each split for  $n = 155$  ( $n = 185$ ) for the CN (MCI) subjects, separately, with results in Table 2.

Table 2: Average Root Mean Prediction Error (RMPE) over 100 repetitions, as obtained from predicted responses from the proposed two-step method. Here,  $n_{\text{train}}$  and  $n_{\text{test}}$  denote the sample sizes for the split training and testing data for CN and MCI subjects.

	$n_{\text{train}}$	$n_{\text{test}}$	First Quartile	Mean	Median	Third Quartile.
CN	100	55	0.134	0.204	0.194	0.266
MCI	120	65	0.139	0.199	0.202	0.271

## 5.2 Human mortality data: Remaining life distributions as object responses

We also analyzed lifetables reflecting remaining life distributions human mortality across 28 countries correspond to distributional responses, coupled with various country-specific covariates with the proposed random effects model. Details can be found in the Supplement.

## 6 Concluding remarks

We present a novel random/mixed effects modeling framework for longitudinal/repeated measurements data when data are random objects that reside in a geodesic metric space. The model is an extension of classical random effects models. The basic linearity assumptions in the Euclidean setting become geodesic assumptions for object data; linearity emerges as a special case.

The proposed model and approach has two components. The first component is concerned with modeling and implementing the relation between sparse observations and the underlying geodesics and reflects the subject-specific random effects, in analogy to the classical model, where the intercept and slope of random regression lines constitute the random effects. The second component characterizes the fixed effects that are common to all subjects by incorporating information from external covariates. The connection between the intrinsic geometry of the underlying metric space and con-

ditional Fréchet means implemented through Fréchet regression is the key to achieve interpretable estimation with asymptotic convergence guarantees.

## A. Technical assumptions on second-step Fréchet regression

In Section 3 the final estimates obtained from regressing object responses  $(\hat{m}_{i\oplus}(0), \hat{m}_{i\oplus}(1)) \in (\mathcal{M} \times \mathcal{M}, d_{\mathcal{M}})$  on the Euclidean predictor  $Z_i \in \mathcal{S} \subset \mathbb{R}^p$ ,  $p \geq 1$ ,  $i = 1, \dots, n$ , using model (9), are  $\hat{\zeta}_{\oplus}(z) = (\hat{\zeta}_0(z), \hat{\zeta}_1(z))^{\top}$ , where

$$\hat{\zeta}_k(z) = \operatorname{argmin}_{\mu \in \mathcal{M}} M_n^{(k)}(\mu), \text{ where } M_n^{(k)}(\mu) = \frac{1}{n} \sum_{i=1}^n s_{in}(Z_i, z) d^2(\mu, \hat{m}_{i\oplus}(k)), \quad k = 0, 1, \quad (21)$$

with empirical weights for the GFR estimator as in (14). Define the intermediate targets

$$\tilde{\zeta}_k(z) = \operatorname{argmin}_{\mu \in \mathcal{M}} \tilde{M}_n^{(k)}(\mu), \text{ where } \tilde{M}_n^{(k)}(\mu) = \frac{1}{n} \sum_{i=1}^n s_{in}(Z_i, z) d^2(\mu, m_{i\oplus}(k)), \quad k = 0, 1, \quad (22)$$

where the empirical GFR weights are defined as before. When object responses lie on geodesics without error, the GFR paths recover the underlying geodesic paths and estimates  $\hat{\zeta}_k$  coincide with the  $\tilde{\zeta}_k$  in (22) for  $k = 0, 1$ . Next we list the assumptions required for the theory of GFR (Petersen and Müller, 2019) that we adopt for this estimation step.

(R0) The objects  $\zeta_k(z)$ ,  $\tilde{\zeta}_k(z)$ , and  $\hat{\zeta}_k(z)$ ,  $k = 0, 1$ , exist and are unique, the latter two almost surely and for any  $\varepsilon > 0$ ,

$$\inf_{d(\mu, \zeta_k(z)) > \varepsilon} M^{(k)}(\mu, z) - M^{(k)}(\zeta_k(z), z) > 0, \quad k = 0, 1.$$

(R1) For  $k = 0, 1$ , let  $B_{\delta}(\zeta_k(z))$  be the ball of radius  $\delta$  centered at  $\zeta_k(z)$  and  $N(\varepsilon, B_{\delta}(\zeta_k(z)), d)$  be its covering number using balls of size  $\varepsilon$ . Then

$$\sqrt{1 + \log N(\varepsilon, B_{\delta}(\zeta_k(z)), d)} d\varepsilon = O(1) \text{ as } \delta \rightarrow 0.$$

(R2) There exist  $\tilde{\eta}_k > 0, \tilde{C}_k > 0$ , possibly depending on  $z$ , such that  $d(\mu, \zeta_k(z)) < \tilde{\eta}_k$  implies

$$M^{(k)}(\mu, z) - M^{(k)}(\zeta_k(z), z) \geq \tilde{C}_k d^2(\mu, \zeta_k(z)), \quad k = 0, 1.$$

Assumption (R0) is commonly used to establish the consistency of an M-estimator such as  $m_{i\oplus}(t)$ ; see Chapter 3.2 in Van der Vaart and Wellner (2000). In particular, it ensures that weak convergence of the empirical process  $\tilde{M}_n$  to the population process  $M$  implies convergence of their minimizers. Furthermore, existence follows immediately if  $\mathcal{M}$  is compact. The conditions on the covering number in Assumption (R1) and curvature in Assumption (R2) arise from empirical process theory and control the behavior of  $\tilde{M}_n - M$  near the minimum, which is necessary to obtain rates of convergence.

## Acknowledgements

Data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](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 can be found at: <http://adni.loni.usc.edu/wp-content/uploads/howtoapply/ADNIAcknowledgementList.pdf>. Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012).

## References

- Afsari, B. (2011) Riemannian  $L^p$  center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, **139**, 655–673.
- Ahidar-Coutrix, A., Le Gouic, T. and Paris, Q. (2020) Convergence rates for empirical barycenters in metric spaces: curvature, convexity and extendable geodesics. *Probability Theory and Related Fields*, **177**, 323–368.
- Alexander-Bloch, A., Giedd, J. N. and Bullmore, E. (2013) Imaging structural co-variance between human brain regions. *Nature Reviews Neuroscience*, **14**, 322–336.
- Allasonniere, S., Chevallier, J. and Oudard, S. (2017) Learning spatiotemporal piecewise-geodesic trajectories from longitudinal manifold-valued data. In *Advances in Neural Information Processing Systems* (eds. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett), vol. 30. Curran Associates, Inc.

- Allen, E., Damaraju, E., Plis, S., Erhardt, E., Eichele, T. and Calhoun, V. (2014) Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex*, **24**, 663–676.
- Ambrosio, L. and Gigli, N. (2008) Construction of the parallel transport in the Wasserstein space. *Methods and Applications of Analysis*, **15**, 1–30.
- Amemiya, Y. (1985) Instrumental variable estimator for the nonlinear errors-in-variables model. *Journal of Econometrics*, **28**, 273–289.
- Arsigny, V., Fillard, P., Pennec, X. and Ayache, N. (2007) Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM journal on matrix analysis and applications*, **29**, 328–347.
- Bhattacharya, R. and Patrangenaru, V. (2003) Large sample theory of intrinsic and extrinsic sample means on manifolds. *The Annals of Statistics*, **31**, 1–29.
- (2005) Large sample theory of intrinsic and extrinsic sample means on manifolds:II. *The Annals of Statistics*, **33**, 1225–1259.
- Billera, L. J., Holmes, S. P. and Vogtmann, K. (2001) Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics*, **27**, 733–767.
- Burago, D., Burago, I. D., Burago, Y., Ivanov, S., Ivanov, S. V. and Ivanov, S. A. (2001) *A Course in Metric Geometry*, vol. 33. Providence, RI: American Mathematical Soc.
- Carroll, R. J. and Hall, P. (2004) Low order approximations in deconvolution and regression with errors in variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **66**, 31–46.
- Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006) *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman and Hall/CRC.
- Chen, X., Hong, H. and Nekipelov, D. (2011) Nonlinear models of measurement errors. *Journal of Economic Literature*, **49**, 901–937.
- Chen, Y., Dubey, P., Müller, H.-G., Bruchhage, M., Wang, J.-L. and Deoni, S. (2021) Modeling sparse longitudinal data in early neurodevelopment. *NeuroImage*, **237**, 118079.
- Chen, Y., Gajardo, A., Fan, J., Zhong, Q., Dubey, P., Bhattacharjee, S., Han, K. and Müller, H. (2020) frechet: statistical analysis for random objects and non-euclidean data. *R package version 0.2. 0*.
- 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. *Annals of Statistics*, **50**, 1573–1592.
- Damoiseaux, J. S., Prater, K. E., Miller, B. L. and Greicius, M. D. (2012) Functional connectivity tracks clinical deterioration in Alzheimer’s disease. *Neurobiology of Aging*, **33**, 828–e19.
- Di Marzio, M., Panzera, A. and Taylor, C. C. (2014) Nonparametric regression for spherical data. *Journal of the American Statistical Association*, **109**, 748–763.

- Diggle, P., Diggle, P. J., Heagerty, P., Liang, K.-Y., Zeger, S. et al. (2002) *Analysis of Longitudinal Data*. Oxford university press.
- Dong, Y. and Wu, Y. (2022) Fréchet kernel sliced inverse regression. *Journal of Multivariate Analysis*, **191**, 105032.
- Dryden, I. L., Koloydenko, A. and Zhou, D. (2009) Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Annals of Applied Statistics*, **3**, 1102–1123.
- Dryden, I. L., Pennec, X. and Peyrat, J.-M. (2010) Power euclidean metrics for covariance matrices with application to diffusion tensor imaging. *arXiv preprint arXiv:1009.3045*.
- Dubey, P. and Müller, H.-G. (2020) Functional models for time-varying random objects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **82**, 275–327.
- Eltzner, B. and Huckemann, S. F. (2019) A smeary central limit theorem for manifolds with application to high-dimensional spheres. *The Annals of Statistics*, **47**, 3360–3381.
- Fan, J. and Müller, H.-G. (2021) Conditional Wasserstein barycenters and interpolation/extrapolation of distributions. *arXiv preprint arXiv:2107.09218*.
- Fréchet, M. R. (1948) Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'institut Henri Poincaré*, **10**, 215–310.
- Fruchterman, T. M. and Reingold, E. M. (1991) Graph drawing by force-directed placement. *Software: Practice and experience*, **21**, 1129–1164.
- González-Briones, A., Villarrubia, G., Paz, J. F. D. and Corchado, J. (2018) A multi-agent system for the classification of gender and age from images. *Comput. Vis. Image Underst.*, **172**, 98–106.
- Hein, M. (2009) Robust nonparametric regression with metric-space valued output. *Advances in neural information processing systems*, **22**.
- Kueper, J. K., Speechley, M. and Montero-Odasso, M. (2018) The Alzheimer’s disease assessment scale–cognitive subscale (adas-cog): modifications and responsiveness in pre-dementia populations. a narrative review. *Journal of Alzheimer’s Disease*, **63**, 423–444.
- Laird, N. M. and Ware, J. H. (1982) Random effects models for longitudinal data. *Biometrics*, **38**, 963–974.
- Latora, V. and Marchiori, M. (2001) Efficient behavior of small-world networks. *Physical review letters*, **87**, 198701.
- Li, J., Li, Y. and Hsing, T. (2022) On functional processes with multiple discontinuities. *Journal of the Royal Statistical Society Series B*, **84**, 933–972.
- Lin, Z. (2019) Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition. *SIAM Journal on Matrix Analysis and Applications*, **40**, 1353–1370.
- Lin, Z. and Müller, H.-G. (2021) Total variation regularized Fréchet regression for metric-space valued data. *The Annals of Statistics*, **49**, 3510–3533.

- Matabuena, M., Petersen, A., Vidal, J. C. and Gude, F. (2021) Glucodensities: a new representation of glucose profiles using distributional data analysis. *Statistical Methods in Medical Research*, **30**, 1445–1464.
- Müller, H.-G. (2016) Peter Hall, functional data analysis and random objects. *The Annals of Statistics*, **44**, 1867–1887.
- Newman, M. E. (2006a) Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, **74**, 036104.
- (2006b) Modularity and community structure in networks. *Proceedings of the national academy of sciences*, **103**, 8577–8582.
- Pellagatti, M., Masci, C., Ieva, F. and Paganoni, A. M. (2021) Generalized mixed-effects random forest: A flexible approach to predict university student dropout. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, **14**, 241–257.
- Petersen, A. and Müller, H.-G. (2016) Functional data analysis for density functions by transformation to a Hilbert space. *The Annals of Statistics*, **44**, 183–218.
- (2019) Fréchet regression for random objects with Euclidean predictors. *The Annals of Statistics*, **47**, 691–719.
- Peyré, G. (2009) Manifold models for signals and images. *Computer Vision and Image Understanding*, **113**, 249–260.
- Pigoli, D., Aston, J. A., Dryden, I. L. and Secchi, P. (2014) Distances and inference for covariance operators. *Biometrika*, **101**, 409–422.
- Rice, J. A. and Wu, C. O. (2001) Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, **57**, 253–259.
- Schennach, S. M. (2016) Recent advances in the measurement error literature. *Annual Review of Economics*, **8**, 341–377.
- Schiratti, J.-B., Allasonniere, S., Routier, A., Colliot, O., Durrleman, S., Initiative, A. D. N. et al. (2015) A mixed-effects model with time reparametrization for longitudinal univariate manifold-valued data. In *International Conference on Information Processing in Medical Imaging*, 564–575. Springer.
- Schötz, C. (2022) Nonparametric regression in nonstandard spaces. *Electronic Journal of Statistics*, **16**, 4679–4741.
- Schwarz, A. J. and McGonigle, J. (2011) Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data. *Neuroimage*, **55**, 1132–1146.
- Şentürk, D. and Nguyen, D. V. (2011) Varying coefficient models for sparse noise-contaminated longitudinal data. *Statistica Sinica*, **21**, 1831.
- Small, C. G. (2012) *The Statistical Theory of Shape*. Springer Science & Business Media.
- Staniswalis, J. G. and Lee, J. J. (1998) Nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, **93**, 1403–1418.

- Takatsu, A. (2011) Wasserstein geometry of Gaussian measures. *Osaka Journal of Mathematics*, 1005–1026.
- Tavakoli, S., Pigoli, D., Aston, J. A. and Coleman, J. (2019) A spatial modeling approach for linguistic object data: Analysing dialect sound variations across Great Britain(with discussion). *Journal of the American Statistical Association*, **114**, 1081–1096.
- Tsochantaridis, I., Hofmann, T., Joachims, T. and Altun, Y. (2004) Support vector machine learning for interdependent and structured output spaces. In *Proceedings of the Twenty-first International Conference on Machine Learning*, 104.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B. and Joliot, M. (2002) Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *NeuroImage*, **15**, 273–289.
- Van der Vaart, A. and Wellner, J. (2000) *Weak Convergence and Empirical Processes: with Applications to Statistics (Springer Series in Statistics)*. Springer, corrected edn.
- Verbeke, G., Molenberghs, G. and Rizopoulos, D. (2010) Random effects models for longitudinal data. In *Longitudinal research with latent variables*, 37–96. Springer.
- Wu, L. (2009) *Mixed effects models for complex data*. Chapman and Hall/CRC.
- Yao, F., Lei, E. and Wu, Y. (2015) Effective dimension reduction for sparse functional data. *Biometrika*, **102**, 421–437.
- Yao, F., Müller, H.-G. and Wang, J.-L. (2005) Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100**, 577–590.
- Yue, X., Park, J. G., Liang, Z. and Shi, J. (2020) Tensor mixed effects model with application to nanomanufacturing inspection. *Technometrics*, **62**, 116–129.
- Zhang, H.-Y., Wang, S.-J., Liu, B., Ma, Z.-L., Yang, M., Zhang, Z.-J. and Teng, G.-J. (2010) Resting brain connectivity: changes during the progress of Alzheimer disease. *Radiology*, **256**, 598–606.
- Zhang, Q., Sue, L. and Li, B. (2022) Partial-global Fréchet regression. *Statistica Sinica*, **xx**, xxx–xxx.
- Zhang, Q., Xue, L. and Li, B. (2021) Dimension reduction and data visualization for Fréchet regression. *arXiv preprint arXiv:2110.00467*.
- Zhu, C. and Müller, H.-G. (2023) Autoregressive optimal transport models. *Journal of the Royal Statistical Society Series B*, **xx**, xxx–xxx.

# Supplementary Materials

## S.1. Additional data illustration and simulation results

This section provides further illustrations of data applications and simulations. Random objects considered in the additional data demonstrations discussed in this section are univariate probability distributions with compact support endowed with the Wasserstein-2 metric (applied to human mortality data) and data that reside on the surface of a sphere, endowed with the geodesic distance. Further illustrations of the proposed method include additional plots for the ADNI study, continuing from Section 5.1 of the main manuscript.

### S.1.1. Simulation study: Responses lying on the surface of a sphere

We next implement our methodology when the responses lie on a Riemannian manifold. In particular, we consider responses lying on the surface of a unit sphere  $S^2 \subset \mathbb{R}^3$  with the center being the origin. The geodesic distance between any two points  $\omega_1$  and  $\omega_2$  lying on the surface of the unit sphere  $S^2$  is given by  $d_g(\omega_1, \omega_2) = \arccos(\omega_1^\top \omega_2)$ . We first model the conditional expectation of the end points of the underlying subject-specific geodesic, conditional on the covariates  $Z$ , as

$$\begin{aligned} \mathbb{E}(\nu_{ik} | Z_i = z, T_{ij} = u) &= \xi_{u,z} \\ &= (\sqrt{(1-z^2)} \cos(\pi u), \sqrt{(1-z^2)} \sin(\pi u), z), \quad z \in (0, 1), \quad k = 0, 1, \quad j = 1, n_i. \end{aligned} \quad (23)$$

The above quantifies the true time-varying regression function conditional on the baseline covariates. In order to generate random realizations of the end-points according to model (23), we first sample the time points at which the repeated measurements are made for each subject, denoted by  $T_{ij}$ , according to a sparse or a dense design as before (see Section 4 in the main manuscript) such that  $T_{ij} \in [0, 1]$ . Further, the baseline covariates  $Z_i$  are generated i.i.d. from  $Unif(0, 1)$  for  $j = 1, \dots, n_i$ ,  $i = 1, \dots, n$ .

The true responses on the surface of the sphere  $S^2$  at the two end points of the underlying geodesic corresponding to the  $i^{\text{th}}$  individual, for  $i = 1, \dots, n$ , are then constructed as follows. A bivariate noise random vector is generated on the tangent space  $T_{\gamma_{\nu_{i0}, \nu_{i1}}(u, z)}^{(i)}(\Omega)$ . To this end, we

define, for  $j = 1, n_i$ ,  $\psi_{ij} = \arcsin(T_{ij})$  and  $\theta_{ij} = \pi T_{ij}$ . An orthonormal basis for the tangent space is denoted by  $(b_{ij}^{(1)}, b_{ij}^{(2)})$ , where  $b_{ij}^{(1)} = (\cos(\psi_{ij}) \cos(\theta_{ij}), \cos(\psi_{ij}) \sin(\theta_{ij}), -\sin(\psi_{ij}))^\top$  and  $b_{ij}^{(2)} = (\sin(\theta_{ij}), -\cos(\theta_{ij}), 0)^\top$ . Adding a noise level  $\sigma^2 = 0.2$ , bivariate random vectors  $A_{ij} = c_{i1} b_{ij}^{(1)} + c_{i2} b_{ij}^{(2)}$  are computed, where  $C_i = (c_{i1}, c_{i2})^\top \stackrel{i.i.d.}{\sim} N_2(0, \sigma^2 I_2)$ . Finally, the responses are generated as

$$\nu_{ik} = \cos(\|A_{ij}\|_E) \zeta_{T_{ij}, Z_i} + \sin(\|A_{ij}\|_E) \frac{A_{ij}}{\|A_{ij}\|_E}, \quad j = 1, n_i, \quad k = 0, 1,$$

with  $\|\cdot\|_E$  being the Euclidean norm. The simulation steps above produce a point  $\nu_{ik}$  on the surface of the two-dimensional sphere at the endpoints  $k = 0, 1$  of some underlying geodesic paths on the surface of the sphere. To complete this step, the geodesic path connecting  $\nu_{i0}$  and  $\nu_{i1}$  is given by  $t \mapsto \frac{1}{\sin \omega} [\nu_{i0} \sin((1-t)\omega) + \nu_{i1} \sin(t\omega)]$ ,  $t \in [0, 1]$ , where  $\omega = \arccos(\nu_{i0}^\top \nu_{i1})$ .

Now, the observable noisy responses are obtained by adding a small perturbation to the random end-points on the geodesic. To this end, we represent any point on the surface of the sphere in spherical coordinates and add noise to the angle the point makes with the  $z$ -axis. A point  $P$  on the surface of the sphere given by  $P = (\rho \sin \phi \cos \theta, \rho \sin \phi \sin \theta, \rho \cos \phi)$ , where  $\rho$  is the distance from  $P$  to the origin,  $\theta$  is the angle between the positive  $x$ -axis and the line segment from the origin to the projection of  $P$  to the  $xy$ -plane, and  $\phi$  is the angle between the positive  $z$ -axis and the line segment from the origin to  $P$ . A noisy observation around  $P$  with a perturbation level  $\alpha_n$  is generated as  $P' := (\rho \sin(\phi + \varepsilon) \cos \theta, \rho \sin(\phi + \varepsilon) \sin \theta, \rho \cos(\phi + \varepsilon))$ , where  $\varepsilon = \pm \alpha_n$  with equal probability  $1/2$ . For this perturbation scheme, the perturbed point  $P'$  has norm  $\rho^2$ , i.e.,  $P'$  still lies on the surface of the sphere  $S^2$ . Further,  $\mathbb{E}(d_g^2(P, P')) = \arccos(P^\top P' / \rho^2) = \alpha_n^2 \rightarrow 0$  as  $n \rightarrow \infty$ . Thus, using the polar coordinate representation of every point generated on the geodesic, the noisy observations are procured as described above.

The simulation study is then carried out for different sample sizes  $n = 50, 400$ , and  $1000$ , for both sparse and dense designs, while fixing the noise level at  $\alpha = 0.1$ . A measure of the efficacy for the fits is constructed as the Integrated Squared Error (ISE) over 500 Monte Carlo simulation runs as

$$\text{ISE}_r = \int_{z \in \mathcal{S}} \int_{t \in [0, 1]} d_g(Y^r(t, z), \hat{Y}^r(t, z)) dt dz, \quad (24)$$

where  $Y^r(t, z)$  and  $\hat{Y}^r(t, z)$  denote, respectively, the true object on the two-dimensional sphere, lying on a geodesic (without perturbation), and the estimated object at time point  $t$  and covariate value  $z$  for the  $r^{\text{th}}$  simulation run, where  $r = 1, \dots, 500$ . Here  $d_g$  denotes the geodesic distance between two points on a sphere and is given by

$$d_g(A, B) = \arccos(A^\top B).$$

where  $A$  and  $B$  are two points on the surface of a sphere. Figure 7 shows that with a denser design

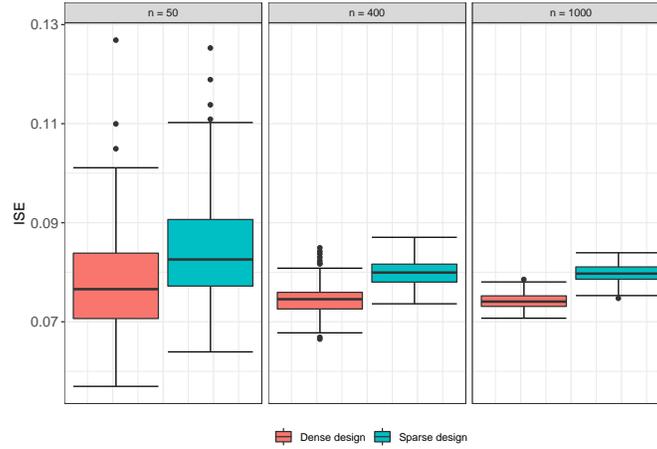


Figure 7: Boxplots of Mean Integrated Squared Errors (MISE) calculated as per (24), over 500 simulation runs and different sample sizes for object responses situated on the surface of a 2-dimensional sphere, corresponding to setting I.

and higher sample size, the ISE reduces significantly, giving evidence for the asymptotic convergence of the estimates to the true underlying object responses.

### S.1.2. Data analysis: Remaining life distributions as object responses

The Human Mortality Database (<https://www.mortality.org/>) provides yearly life table data for males and females and various countries. Here we study the time-varying association between remaining life distribution and various socioeconomic indices at the country level and consider the life tables for females over 30 calendar years, 1990 – 2019, for  $n = 28$  countries. We consider the remaining life distribution  $R(t) = P(T \leq t | T \geq 75)$  as responses, where  $T$  denotes age-at-death and the remaining life distribution is considered on the interval  $[75, 120]$  (all in years). This remaining life distribution and its density can be easily obtained from the available lifetable data

that correspond to histograms with bin width one year by adding a smoothing step, for which we used the R package *frechet* (Chen et al., 2020) with bandwidth 2 years.

We then obtained a sample of time-varying univariate probability distributions which are the responses for  $n = 28$  countries, where the time axis represents 30 calendar years from  $T := [1990, 2019]$  and the observation made at each calendar year for each country corresponds to the remaining life distribution over the age interval  $[75, 120]$ .

For the first-step regression, we fit model (12) to obtain the estimates for the remaining life distribution at the first (time 0, corresponding to the year 1990) and last (time 1, corresponding to the year 2019) point of our time domain. The inherent assumption is that, for each country, the remaining life distributions over the years are observed around some geodesic in the Wasserstein-2 space of distributions with small error/perturbation, where the underlying geodesic connects the two distribution objects corresponding to time 0 and time 1. The fitted responses  $\hat{m}_{i\oplus}(0)$  and  $\hat{m}_{i\oplus}(1)$  are then treated as a summary of the time-varying remaining life distributions for the  $i^{\text{th}}$  country,  $i = 1, \dots, 28$ , and are carried forward as the paired distributional response to the second-step regression as per (13).

For implementing the second step regression as per model (8), we consider a 4– dimensional baseline covariate for each country, where the covariates for the  $i^{\text{th}}$  country represent (1) Unemployment rate (% of the total labor force) (2) Fertility Rate (Births per women), (3) GDP per capita- International purchasing power parity, and (4) Population growth (annual %), measured in the calendar year 1990. The data is obtained from the World Bank Database at <https://data.worldbank.org>. Our aim is to quantify the effects of this baseline/ external covariate, possibly changing over the calendar years, on the remaining life distributions. The second-step regression with the paired object responses  $(\hat{m}_{i\oplus}(0), \hat{m}_{i\oplus}(1))$  and Euclidean covariates  $Z_i, i = 1 \dots, n$ , produces the fitted objects  $\hat{\zeta}_{\oplus}(z) = (\hat{\zeta}_0(z), \hat{\zeta}_1(z))^{\top}$  over varying values  $Z = z$ .

It is of interest to see how the estimated distributions at times 0 and 1 given by  $\hat{\zeta}_0(z)$  and  $\hat{\zeta}_1(z)$ , respectively, change over varying levels of the baseline covariate  $Z$ . Here  $Z$  is a 4-dimensional predictor. To elicit the effect of each component of  $Z$ , we vary the levels of that component from low to high while keeping the other three components fixed at their mean level. For example, Figure 8 illustrates how the remaining-life density changes with increasing levels of GDP per capita, while the other three predictors are kept fixed at their mean levels. The left and right panels display the

fitted densities for the calendar years 1990 and 2019 respectively. The fitted densities are color coded such that blue to red indicates smaller to larger value of GDP. We find that smaller values of GDP are associated with left-shifted remaining life distributions, while a larger GDP value corresponds to a shift of the mode of the age-at-death toward the right. Further, the densities for the year 1990 are more left-skewed than the ones for 2019, indicating an increasing right shift of the remaining life distribution as calendar time progresses. The time effect and GDP effect are seen to be not simply additive but the GDP effect is more pronounced in 2019 than in 2010.

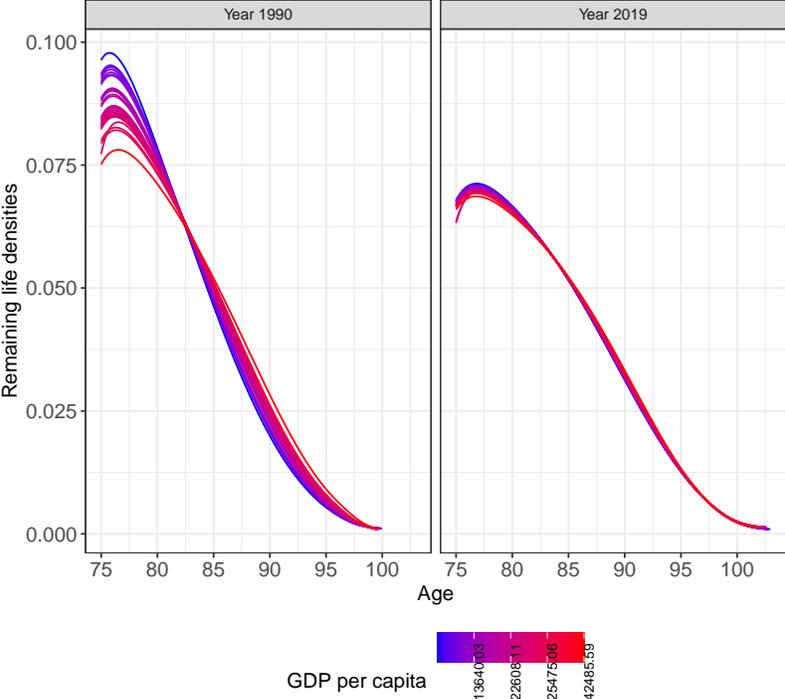


Figure 8: Effect of the covariate GDP per Capita, at the beginning and end of the time domain. The changes in density of the remaining life distribution after age 75 as GDP per Capita rate ranges from low (blue) to high (red) are displayed when the other predictors are fixed at their mean level. The left and right panels show the fits at the calendar years 1990 and 2019, respectively.

For increasing levels of the fertility rate, unemployment rate, and population growth percentage, similar patterns for the time-varying effect of these covariates are observed, but to a lesser extent (See Figures 9, 10, and 11, respectively).

For each country, the fitted geodesics in the Wasserstein space of distributions summarize the time dynamics of the remaining life distributions along with the effects of the covariates. We further demonstrate the interpretability of the proposed random effects model by displaying the fits at the

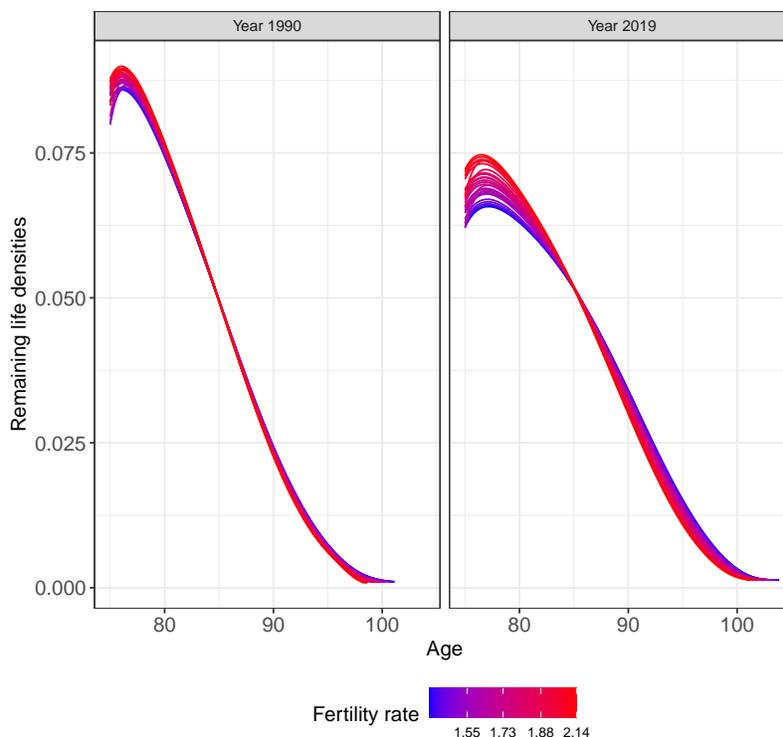


Figure 9: Effect of the covariate Fertility rate, at the beginning and end of the time domain. The changes in density of the remaining life distribution after age 75 as Fertility rate ranges from low (blue) to high (red) are displayed when the other predictors are fixed at their mean level. The left and right panels show the fits at the calendar years 1990 and 2019, respectively.

beginning and end of the time domain when varying the value of one predictor at the 10%, 50%, and 90% quantile levels, while keeping the other two predictors fixed at their mean. We then compute the estimated densities situated on the fitted geodesic in the distribution space corresponding to a grid of time points in [1990, 2019]. The left, middle, and right panels of Figure 12 display the estimated densities at the calendar years 1995, 2000, and 2008, respectively. For each panel, the red, blue, and green lines correspond to the 10%, 50%, and 90% quantile values for GDP per Capita, while the other three predictors are kept fixed at their mean levels. We observe a shift in the remaining life densities towards the right over the years, thus indicating improved remaining survival as calendar time progresses.

Similar interpretations emerge for the other three predictors from the patterns displayed in Figure 13, 14, and 15. We observe that a higher value of the covariate levels is generally associated with right-shifted remaining life distribution,

To summarize, in Figure 16, we illustrate the observed densities for the remaining life distri-

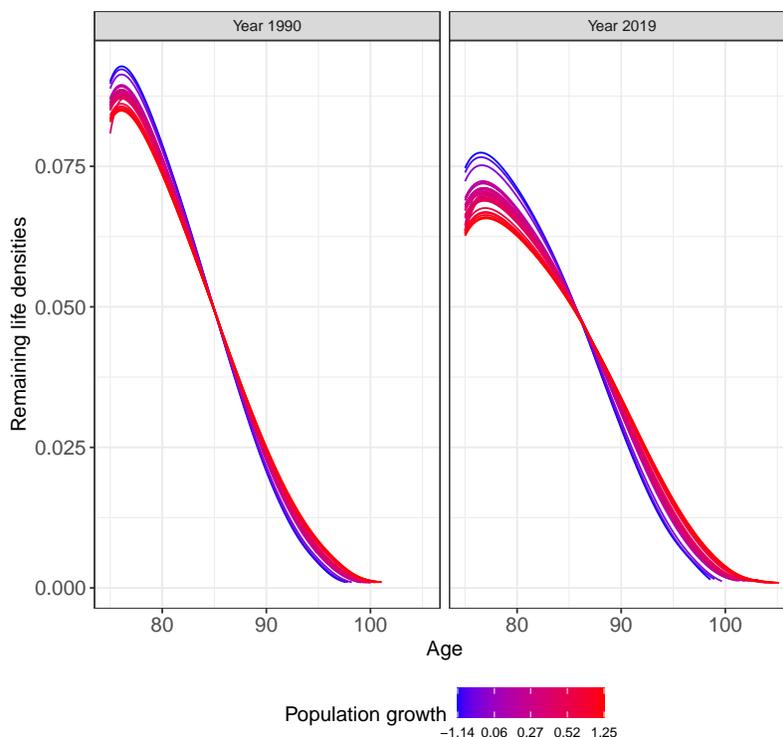


Figure 10: Effect of the covariate Percentage of Population Growth, at the beginning and end of the time domain. The changes in density of the remaining life distribution after age 75 as Percentage of Population Growth ranges from low (blue) to high (red) are displayed when the other predictors are fixed at their mean level. The left and right panels show the fits at the calendar years 1990 and 2019, respectively.

Contributions for a few selected countries over three selected calendar years, along with the densities predicted at the observed baseline-covariate values for that country. The six panels, clockwise from top-left, correspond to Australia, Finland, France, United States, Netherlands, and Japan; while red, green, and blue colors indicate the calendar years 1995, 2000, and 2008, respectively. The observed and predicted densities are plotted in solid and dashed lines for each country and each calendar year, and follow the same temporal pattern. The fits are close to the observations, thus giving evidence for the validity of the model. The small discrepancies in the estimated-vs-observed densities towards the beginning and end of the domain could be caused by boundary effects of the regression fits.

Finally, the performance of the fits is measured by the out-of-sample prediction performance of the proposed method. For this, we first randomly split the dataset into a training set with sample size  $n_{\text{train}} = 18$  and a test set with the remaining  $n_{\text{test}} = 10$  subjects (countries). We then consider the fitted objects  $\hat{\zeta}_{\oplus}(\cdot)$  obtained from the training set and predict the responses in the test set using

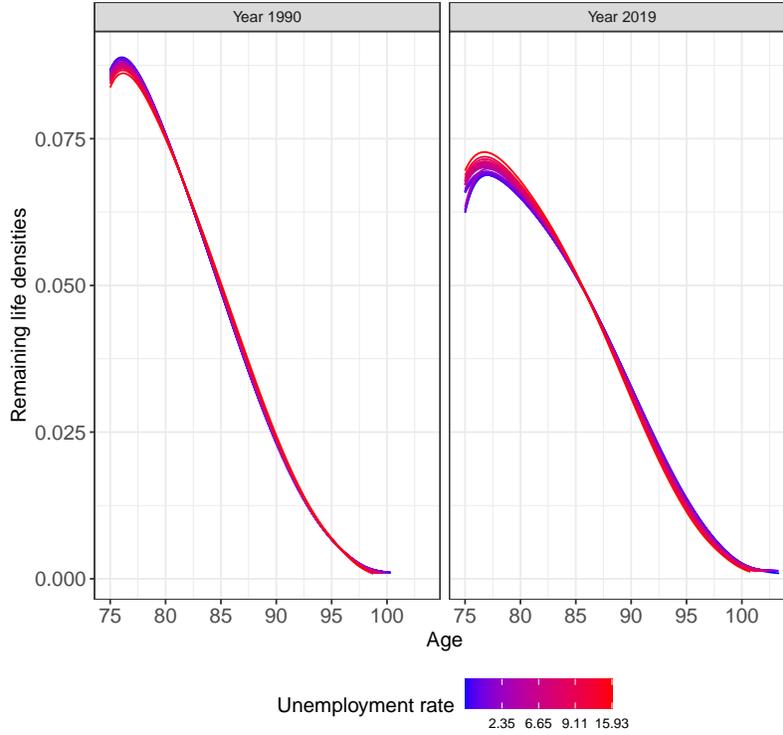


Figure 11: Effect of the covariate Unemployment Rate, at the beginning and end of the time domain. The changes in density of the remaining life distribution after age 75 as Unemployment Rate ranges from low (blue) to high (red) are displayed when the other predictors are fixed at their mean level. The left and right panels show the fits at the calendar years 1990 and 2019, respectively.

the covariates present in the test set. As a measure of the efficacy of the fitted model, we compute the root mean squared prediction error as

$$\text{RMPE} = \left[ \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \frac{1}{n_i} \sum_{j=1}^{n_i} d_W^2 \left( Y_{ij}^{\text{test}}, \hat{Y}_{ij}^{\text{test}} \right) \right]^{-1/2}, \quad (25)$$

where  $Y_{ij}^{\text{test}} = Y_{ij}^{\text{test}}(T_{ij})$  denotes the  $i^{\text{th}}$  observed response at time  $T_{ij}$ ,  $j = 1, \dots, n_i$ , for the  $i^{\text{th}}$  subject in the test set,  $\hat{Y}_{ij}^{\text{test}} = \hat{Y}_{ij}^{\text{test}}(T_{ij})$  denotes predicted object for the second-step fits at the covariate level  $Z_i$  for the predictors in the test set.  $d_W$  denotes the Wasserstein-2 metric in the space of distribution objects. We repeat this process 500 times and compute RMPE for each split for  $n = 28$  countries. separately. The summary of the RMPE is shown in Table 3.

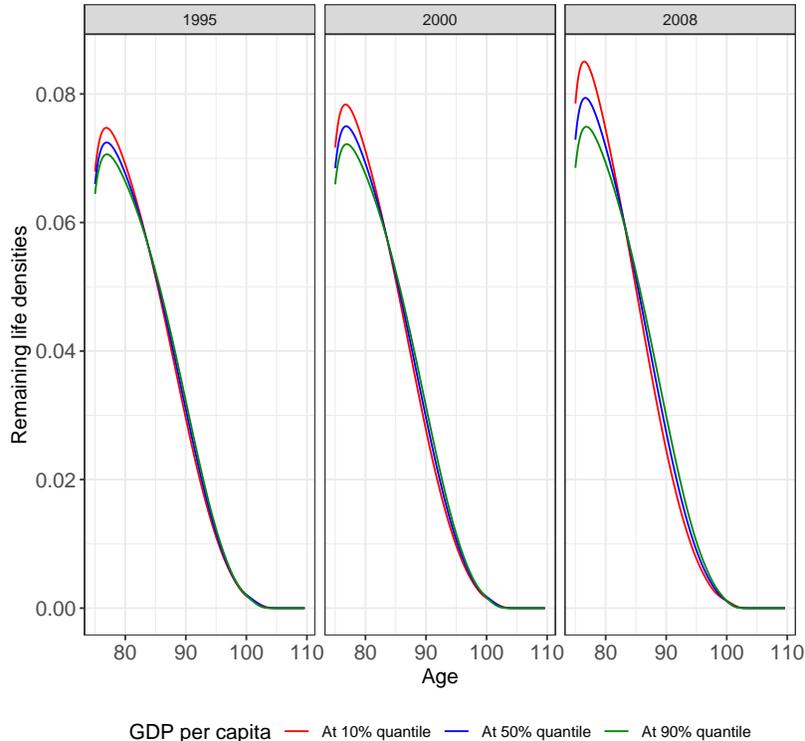


Figure 12: Effect of the predictor GDP per Capita, evaluated at points on the fitted geodesic. The left, middle, and right panels show the fits for the years 1995, 2000, and 2008 respectively, where the remaining life densities are fitted at the 10% (red), 50% (blue) and 90% (green) quantile levels of GDP per Capita, while the other predictors are fixed at their mean levels.

Table 3: Average Root Mean Prediction Error (RMPE) over 500 repetitions, as obtained from predicted responses from the proposed two-step method. Here,  $n_{\text{train}}$  and  $n_{\text{test}}$  denote the sample sizes for the split training and testing datasets respectively.

$n_{\text{train}}$	$n_{\text{test}}$	First Quartile	Mean	Median	Third Quartile.
18	10	0.2418	0.3196	0.2935	0.3656

### S.1.3. ADNI data

Continuing from Section 5.1 in the main manuscript, we illustrate the network structure of the fitted Pearson correlation connectivity (PCC) matrices for CN and MCI subjects. The PCC matrices serve as responses residing in the space of correlation matrices equipped with the power Euclidean metric with power  $\alpha = 1/2$ , coupled with baseline covariates taken as age and C-score over a time window  $[0, 1]$ , since the first available scan.

First, the effect of the C-score for a fixed age is demonstrated for MCI subjects through correlation plots of the estimated PCC matrices. We fixed the age of the subjects at their mean level and fitted the model at varying levels of the C-score, namely, at the 10%, 50%, and 90% quantiles of the

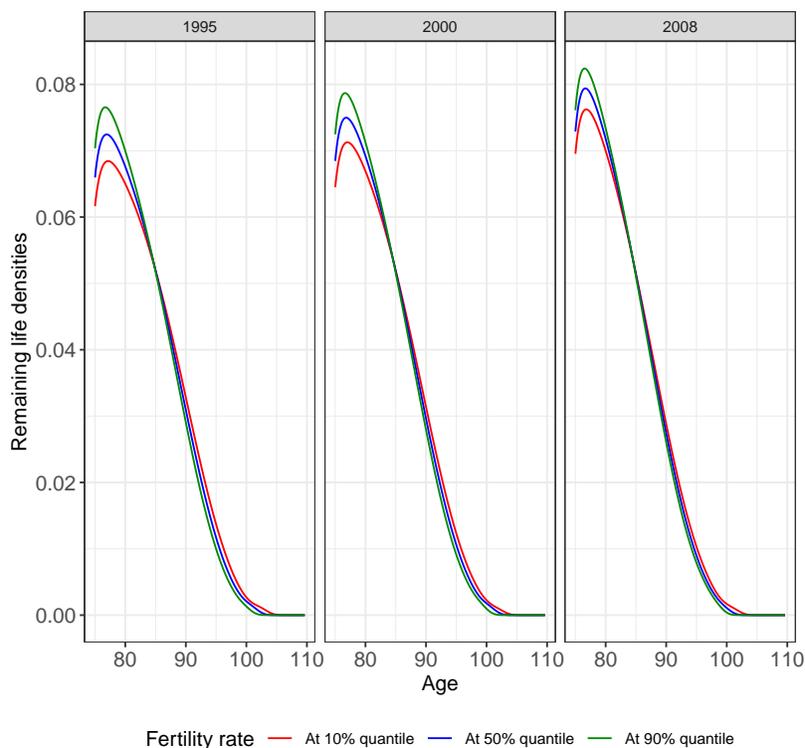


Figure 13: Effect of the predictor Fertility Rate, evaluated at points on the fitted geodesic. The left, middle, and right panels show the fits for the years 1995, 2000, and 2008 respectively, where the remaining life densities are fitted at the 10% (red), 50% (blue) and 90% (green) quantile levels of Fertility rate, while the other predictors are fixed at their mean levels.

C-score. Figure 17 demonstrates the trend for the temporal correlations for varying predictor levels at different times of the study. The top and bottom rows correspond to the predicted correlation matrices (with the diagonals set to 0) at times 0 and 1 respectively, while within each row the left, middle, and right panels depict the fits at the 10%, 50%, and 90% quantiles of the C-score with the age fixed at its mean level. The overall correlation strengths decrease as C-scores increase, reflecting the mean effects of the baseline covariates. Further, comparing the rows for each panel, we find overall weaker correlations at time 1 compared with those at time 0.

We also converted the predicted PCC matrices into networks to better interpret and visualize the brain structure. The predicted PCC matrices at varying levels of C-scores and for a fixed age were converted into weighted adjacency matrices and we explored the community detection methods for these network representations for both the CN and MCI subjects. The predicted networks for the CN and MCI subjects are demonstrated in Figure 18 and 19 respectively, where the nodes were placed using the Fruchterman-Reingold layout algorithm (Fruchterman and Reingold, 1991)

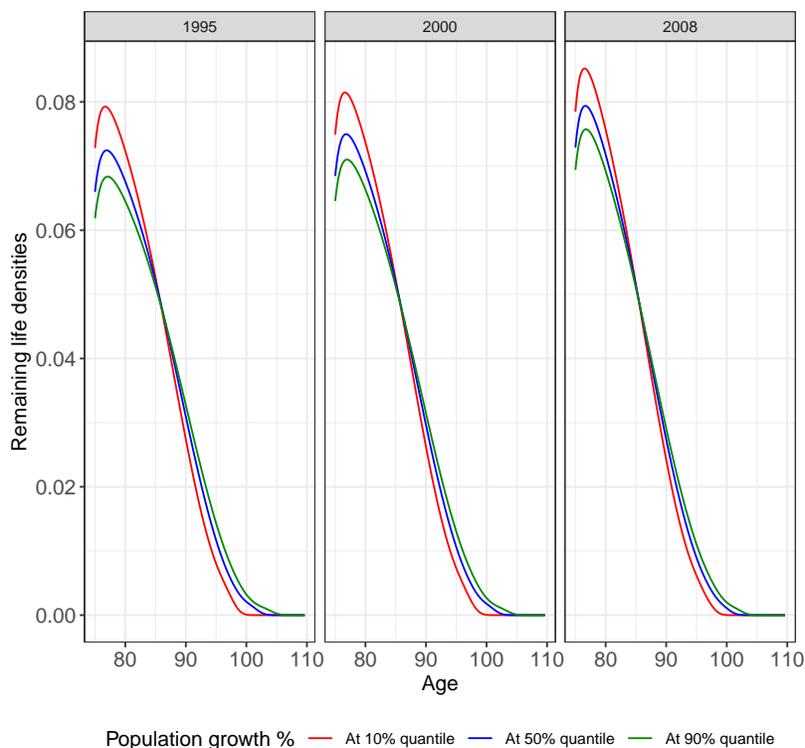


Figure 14: Effect of the predictor Percentage of Population Growth, evaluated at points on the fitted geodesic. The left, middle, and right panels show the fits for the years 1995, 2000, and 2008 respectively, where the remaining life densities are fitted at the 10% (red), 50% (blue) and 90% (green) quantile levels of Population Growth, while the other predictors are fixed at their mean levels.

for visualization. Spectral clustering (Newman, 2006a) is applied to detect the community structure in each network, where different communities are distinguished by different colors. The R package *igraph* was used to find communities in graphs via directly optimizing a modularity score with a fast greedy algorithm.

The number of communities for the CN subjects at the 10%, 50%, and 90% quantiles of the C-score, where the other covariate age is fixed at its mean level, are 7, 6, 7 corresponding to the fits at time 0, respectively, and 13, 12, 11, corresponding to the fits at time 1, respectively. The number of communities found in the predicted networks for MCI subjects are 7, 3, 7 and 11, 4, 11, respectively, at time 0 and time 1. The communities with no less than 10 nodes are highlighted using colored polygons. These communities are found to be associated with different anatomical regions of the brain, where a community is identified as the anatomical region to which the majority of nodes belong. However, the communities found using the spectral clustering method overlap, especially for a higher value of the C-score, as the local interconnectivity and tendency to form a

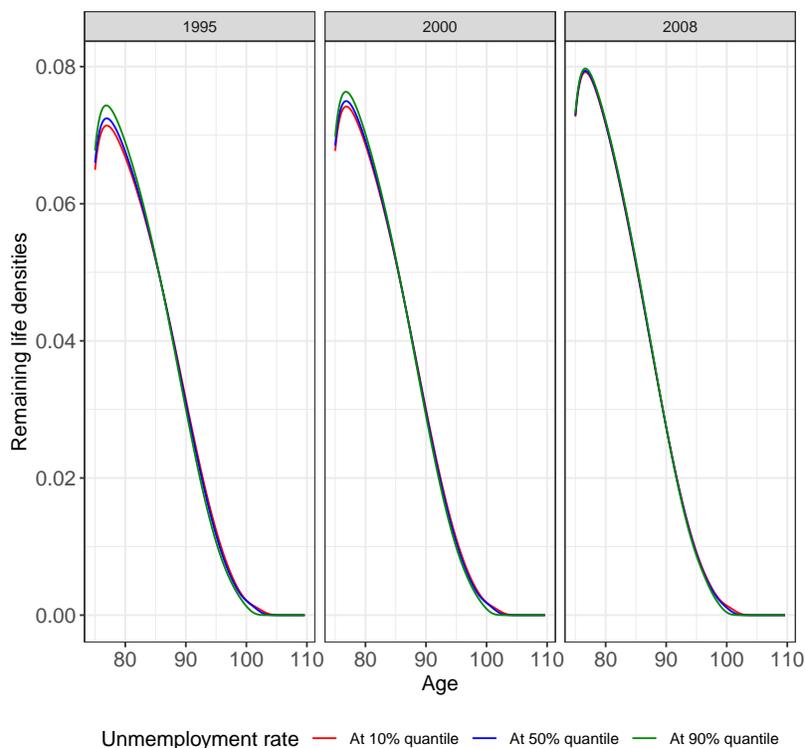


Figure 15: Figure showing the effect of the baseline predictor- Unemployment rate, evaluated at points on the geodesic. The left, middle, and right panels show the fits for the years 1995, 2000, and 2008 respectively, where the remaining life densities are fitted at the 10%, 50%, and 90% quantile levels of the Unemployment rate (shown in red, blue, and green curves, respectively), while the other predictors are fixed at their mean levels.

clique more locally increases. High cliquishness is known to be associated with reduced capability to rapidly combine specialized information from distributed brain regions, which may contribute to the cognitive decline of Alzheimer’s subjects.

Finally, the global efficiency, a characteristic measure of network integration for the estimated networks evaluated at all points between time 0 and 1, on the underlying geodesic in the space of SPD matrices, for the CN and MCI subjects is demonstrated in Figure 20. Global efficiency is a scaled measure of how many steps it takes when moving through the network from one node to another, where higher efficiency means that on average fewer steps are needed (Alexander-Bloch et al., 2013; Latora and Marchiori, 2001). In the left and right panels of Figure 20, the time-varying nature of the global efficiency of the estimated networks are illustrated for the CN and MCI subjects respectively. Each panel shows an overall decreasing trend with time. Further, for each panel, the estimated networks at the 10%, 50%, and 90% quantiles of the C-score are shown in red, blue, and purple, where the other covariate age is kept fixed at its mean level. The purple line is

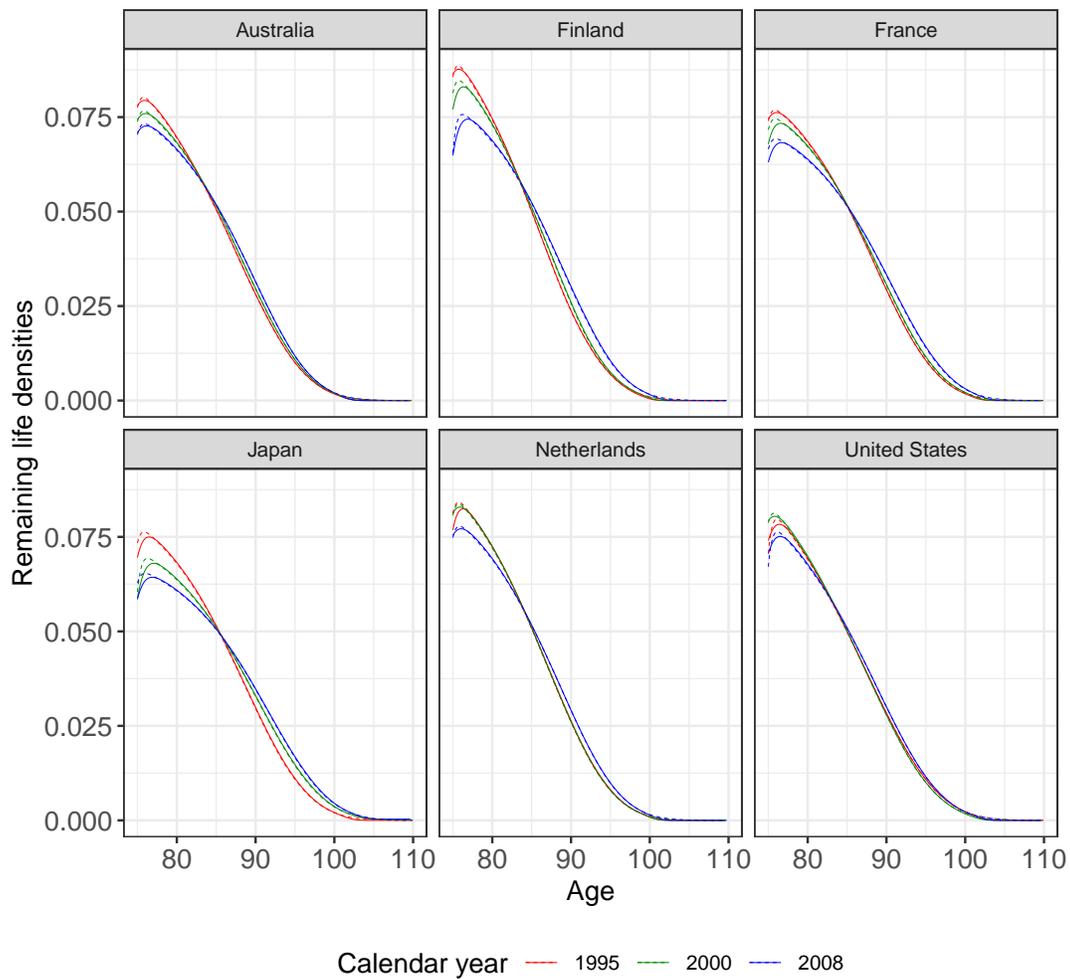


Figure 16: Figure displaying the observed and estimated remaining life distributions, represented as densities for a select few countries over a few selected calendar years. The panels, clockwise from top-left, correspond to the countries- Australia, Finland, France, United States, Netherlands, and Japan. In each panel, the red, blue, and green lines show the densities at the calendar years 1995, 2000, and 2008, respectively. The observed and predicted densities are shown in solid and dashed lines, respectively, the estimated densities being computed at the observed covariate values.

generally below the others, which suggests that higher C-scores are associated with lower degrees of global efficiency, indicating less connectivity in the brain and an enhanced cognitive deficiency. The impairment over time looks more severe for the MCI subjects.

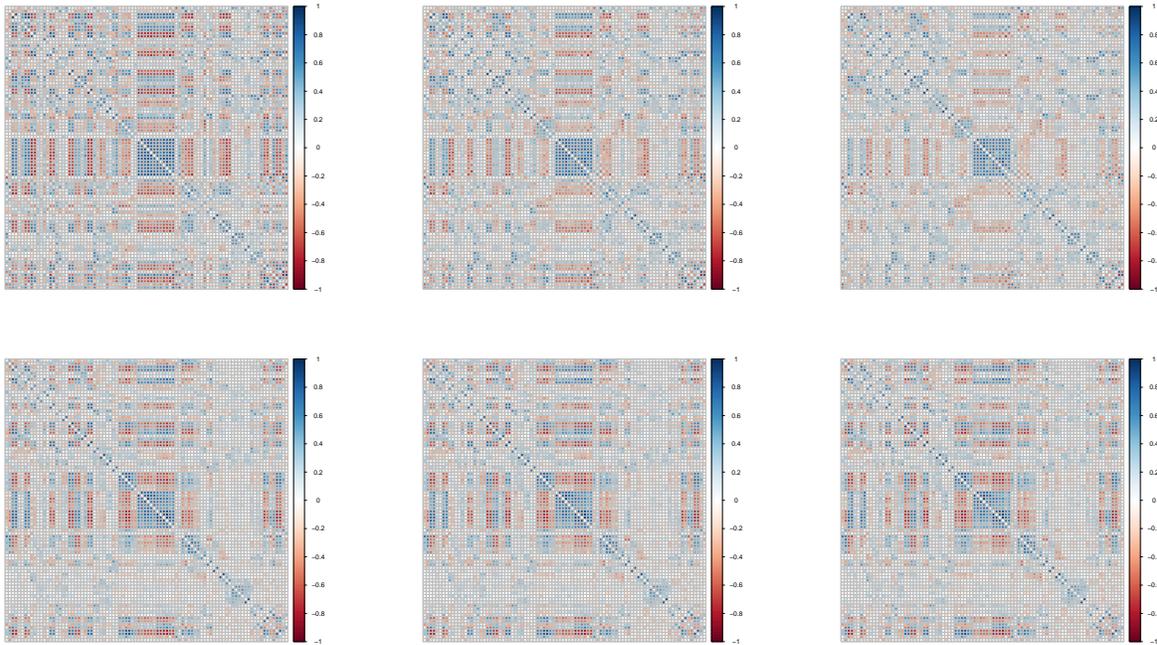


Figure 17: Correlation plot of the predicted PCC matrices over varying levels of covariates at two different time points for MCI subjects. The top and bottom rows correspond to the predicted correlation matrices at times 0 and 1 respectively, while within each row the left, middle, and right panels depict the fits at the 10%, 50%, and 90% quantiles of the C-score with the other covariate age fixed at its mean level. Positive (negative) values for correlations are drawn in red (blue).

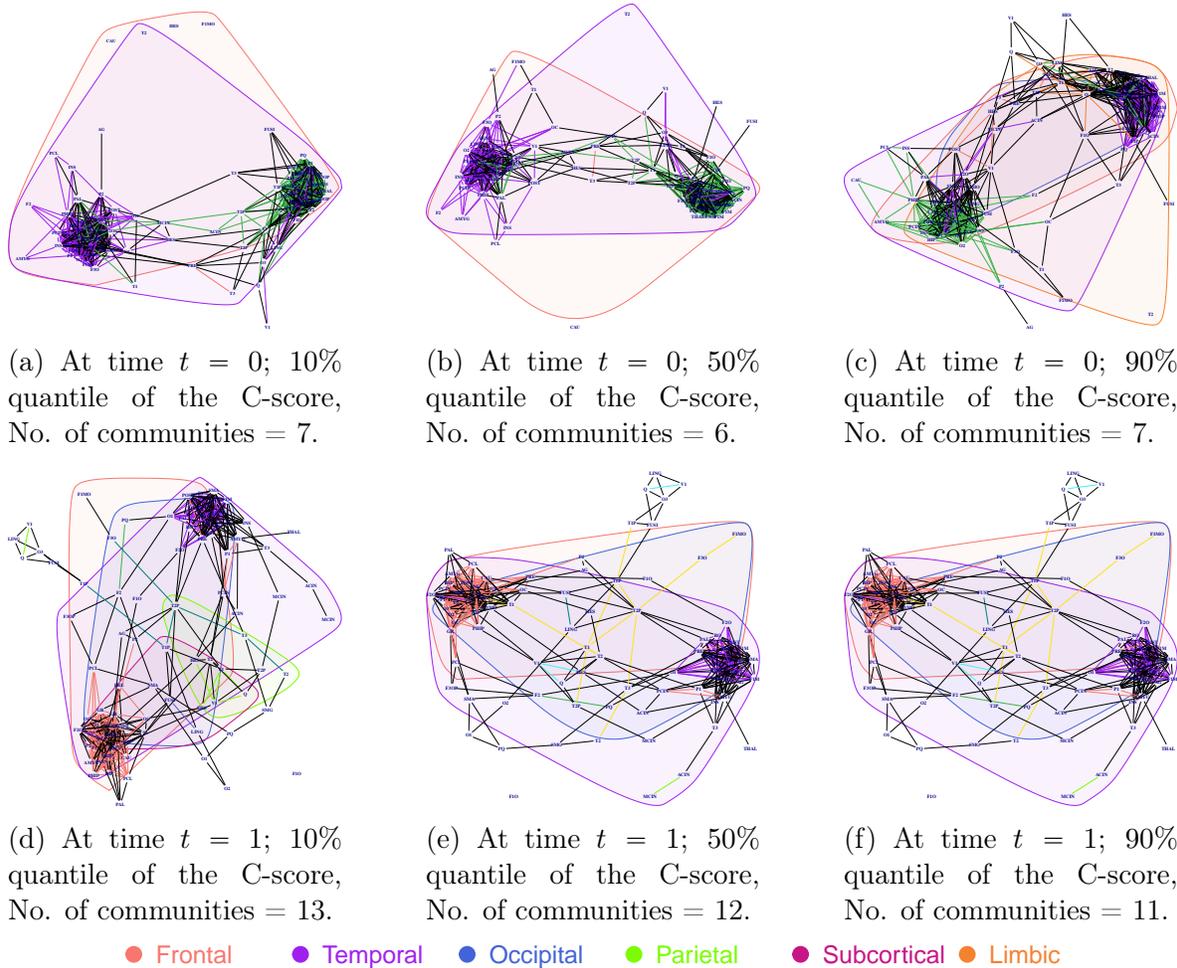
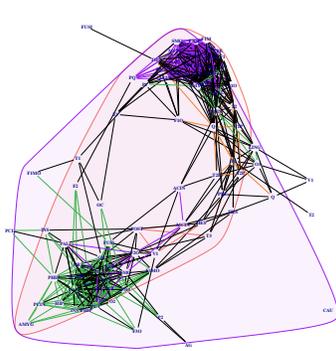
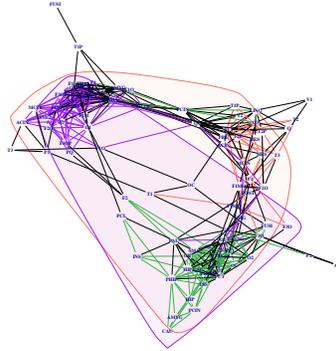


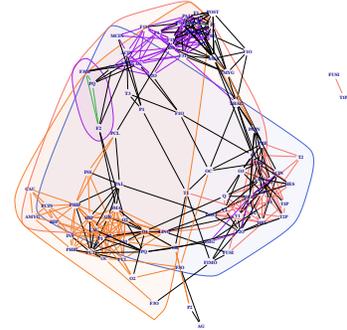
Figure 18: Representation of the estimated PCC matrices at different levels of C-score as brain network, using spectral community detection method for the CN subjects. The top and bottom rows correspond to the predicted correlation matrices at times 0 and 1 respectively, while within each row the left, middle, and right panels depict the fits at the 10%, 50%, and 90% quantiles of the C-score with the other covariate age fixed at its mean level. The communities comprising 10 or more ROIs are highlighted using colored polygons. These communities are found to be associated with different anatomical regions of the brain.



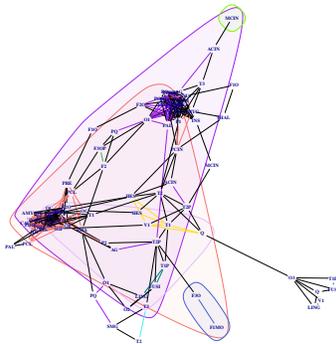
(a) At time  $t = 0$ ; 10% quantile of the C-score, No. of communities = 7.



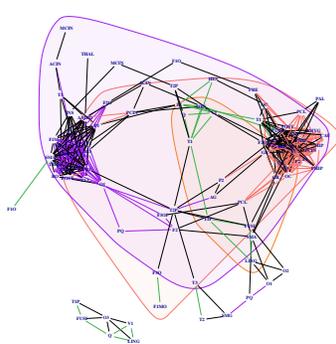
(b) At time  $t = 0$ ; 50% quantile of the C-score, No. of communities = 3.



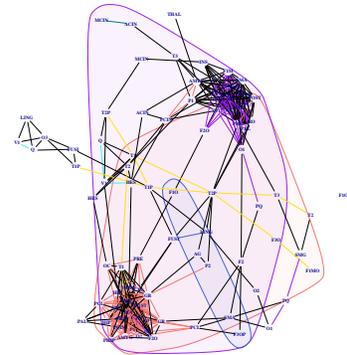
(c) At time  $t = 0$ ; 90% quantile of the C-score, No. of communities = 7.



(d) At time  $t = 1$ ; 10% quantile of the C-score, No. of communities = 11.



(e) At time  $t = 1$ ; 50% quantile of the C-score, No. of communities = 4.



(f) At time  $t = 1$ ; 90% quantile of the C-score, No. of communities = 11.

● Frontal ● Temporal ● Occipital ● Parietal ● Subcortical ● Limbic

Figure 19: Representation of the estimated PCC matrices at different levels of C-score as brain network, using spectral community detection method for the MCI subjects. The top and bottom rows correspond to the predicted correlation matrices at times 0 and 1 respectively, while within each row the left, middle, and right panels depict the fits at the 10%, 50%, and 90% quantiles of the C-score with the other covariate age fixed at its mean level. The communities comprising 10 or more ROIs are highlighted using colored polygons. These communities are found to be associated with different anatomical regions of the brain.

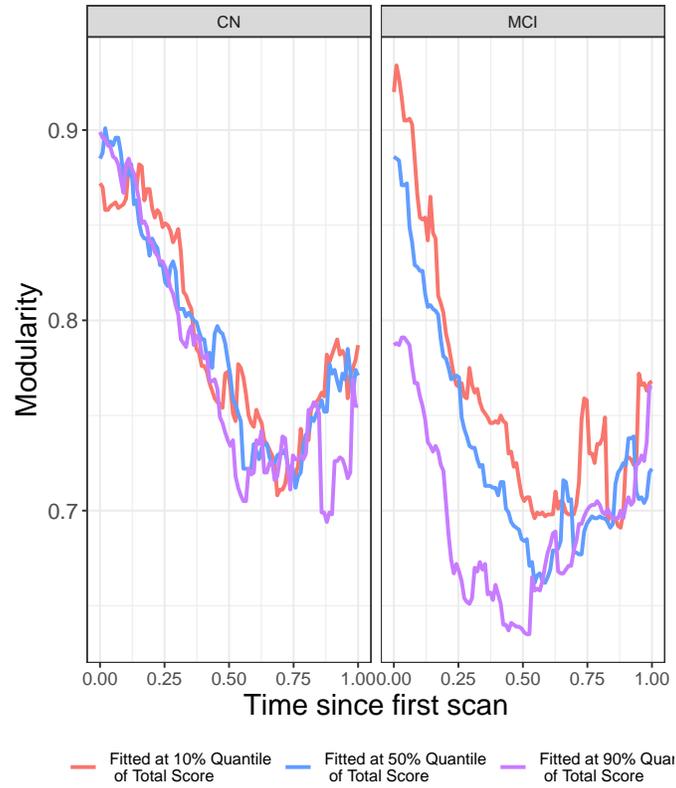


Figure 20: Figure showing the global efficiency of the estimated brain network over time for the CN and MCI subjects (in the left and right panels, respectively). The covariate levels at which the networks are estimated are depicted in red, blue, and purple, respectively, corresponding to the 10%, 50%, and 90% quantiles of the C-score, with the other covariate age fixed at its mean level.