Modeling Sparse Longitudinal Data in Early Neurodevelopment*†

Yaqing Chen[‡1], Paromita Dubey[‡1], Hans-Georg Müller[‡1], Muriel Bruchhage[2,3], Jane-Ling Wang[1], Sean Deoni[§2,3,4,5], and RESONANCE Consortium

[1]Department of Statistics, University of California, Davis, Davis, CA, 95616, USA
[2]Advanced Baby Imaging Lab, Hasbro Children's Hospital, Rhode Island Hospital, Providence, RI, 02903, USA
[3]Department of Pediatrics, Warren Alpert Medical School at Brown University, Providence, RI, 02912, USA
[4]Department of Radiology, Warren Alpert Medical School at Brown University, Providence, RI, 02912, USA
[5]Maternal, Newborn, and Child Health Discovery & Tools, Bill & Melinda Gates Foundation, Seattle, WA, USA

Abstract

Early childhood is a period marked by rapid brain growth accompanied by cognitive and motor development. However, it remains unclear how early developmental skills relate to neuroanatomical growth across time with no growth quantile trajectories of typical brain development currently available to place and compare individual neuroanatomical development. Even though longitudinal neuroimaging data have become more common, they are often sparse, making dynamic analyses at subject level a challenging task. Using the Principal Analysis through Conditional Expectation (PACE) approach geared towards sparse longitudinal data, we investigate the evolution of gray matter, white matter and cerebrospinal fluid volumes in a cohort of 446 children between the ages of 1 and 120 months. For each child, we calculate their dynamic age-varying association between the growing brain and scores that assess cognitive functioning, applying the functional varying coefficient model. Using local Fréchet regression, we construct age-varying growth percentiles to reveal the evolution of brain development across the population. To further demonstrate its utility, we apply PACE to predict individual trajectories of brain development.

Keywords: Brain-for-age growth chart; Brain development percentiles; Concurrent regression modeling; Functional principal components; Individual developmental trajectory reconstruction; Whole brain MRI.

*Declaration of interest: Sean Deoni receives salary and grant support for Nestle. S.A.

†Abbreviations: cdf: cumulative distribution function; FDA: functional data analysis; FPCA: functional principal component analysis; FPC: functional principal component; PACE: Principal Analysis through Conditional Expectation; pGM: proportion of grey matter; pWM: proportion of white matter; pCSF: proportion of cerebrospinal fluid.

‡Co-first authors.

§Corresponding author. Email address: sean_c_deoni@brown.edu

1 Introduction

Infancy and early childhood are periods of rapid physical growth, skill and brain development. Throughout the first year of life, the brain grows from 25% to 75% of adult volume during healthy development, and reaches 95% of its peak size by age six (Giedd and Rapoport, 2010). This increase in brain volume reflects underlying macro- and micro-structure tissue maturation, including increasing myelination and white matter volume, changing cortical morphometry, and increasing sub-cortical gray matter volumes, synapse and neuronal density.

By scanning and assessing children throughout early development, longitudinal studies are in theory able to characterise brain growth patterns consistently with age, and investigate associations with current and future cognitive performance. However, in reality, longitudinal data are often sparse and collected at different time points, as participants miss scans or assessments because of illness or, more recently, self-isolation. In addition, some study designs, such as hybrid or accelerated-longitudinal designs, intentionally scan children across different age ranges in order to quickly collect data across a larger effective age range. The resulting sparse and unbalanced nature of the data makes modeling the time-varying evolution of brain growth patterns a challenge.

Growth patterns are typically studied in form of population-based growth trajectories, making it easier to evaluate current and future development across differing geographies and environmental settings. For example, based on physical growth data, an estimated 165 million children under 5 years of age are currently stunted and thus at risk or are failing to achieve their developmental potential (UNICEF et al., 2012). However, growth curves are mainly used to assess purely physical growth (i.e. length, height, weight), neglecting regions that have been more closely linked to cognitive and motor development, such as the brain (Silbereis et al., 2016). Following WHO guidelines to monitor physical growth (WHO, 2006), Lambda-Mu-Sigma (LMS) and Box-Cox Power Exponential (BCPE) methods (Cole, 1988, 1994; Rigby and Stasinopoulos, 2004) are used to create physical growth curves. LMS and BCPE assume that the age-specific Box-Cox transformation of the original measurements $Y(t)$ follows Gaussian and power exponential distributions, where Gaussian distributions

are a special case of the latter. These distributional assumptions are restrictive and may not hold in practice. For example, the LMS and BCPE methods are restricted to unimodal distributions, while brain development can be characterized by multimodality in age-varying distributions. Without such distributional assumptions, Cox and Jones (in separate contributions to the discussion of Cole, 1988) proposed a nonparametric model to estimate the τth conditional quantiles $g_\tau(\cdot)$ by

$$\widehat{g}_\tau = \underset{g}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau(Y_i - g(X_i)) + \lambda \int g''(z)^2 \mathrm{d}z, \tag{1}$$

where $\rho_\tau(z) = \tau \max\{z, 0\} + (1-\tau)\max\{-z, 0\}$ for $z \in \mathbb{R}$.

Cox's model does not include the regularization, as it sets $\lambda = 0$, while Jones considers the model with general $\lambda \geqslant 0$. Koenker and Bassett (1978) impose a linearity condition $g(x) = x^\top \beta$. However, this method suffers from quantile crossing, i.e. quantile lines $\widehat{g}_\tau$ may (and often do) cross each other for different values of τ (He, 1997). This means that a lower-level quantile could be larger than a higher-level quantile, e.g., a median at certain time can be greater than the third quartile at the same time, which is unrealistic. Various modifications have been proposed to overcome this crossing problem, usually under linearity conditions (Koenker and Bassett, 1978); see also He (1997); Bondell et al. (2010).

Such linearity assumptions may not be valid for real-world data. Structural brain development follows a nonlinear trajectory at both whole-brain and regional brain structure levels (Gennatas et al., 2017; Bray et al., 2015; Gogtay and Thompson, 2010; Lebel et al., 2008; Lebel and Beaulieu, 2011; Giorgio et al., 2010). Gray matter volume increases rapidly during infancy, peaking within the first three years of life (Matsuzawa et al., 2001) and gradually decreases thereafter. In contrast, white matter volume increases throughout childhood and early adolescence (Barnea-Goraly et al., 2005; Blakemore and Choudhury, 2006) before decreasing in older adulthood. Various studies have shown differences in these patterns by biological sex, with boys showing a greater gray matter percentage overall and girls displaying a greater white matter percentage (Giedd and Rapoport, 2010). This leaves a need for methodologies that help characterise longitudinal patterns of brain development, and to develop population-based growth curves of brain development, aiming to investigate individual

variability and benchmark potentially aberrant development. To capture and characterize these nonlinear patterns of development, linear and linear mixed effects models previously used in cross-sectional and longitudinal studies are ill-posed, and more flexible nonparametric methods are needed. Nonparametric methods for conditional quantile estimation have been proposed based on kernel smoothing and splines (e.g., Samanta, 1989; Hendricks and Koenker, 1992), as well as variants of model (1) with different penalties proposed in Koenker et al. (1994). However, these methods either suffer from boundary effects detracting from the global convergence of the estimators (e.g., Müller and Stadtmüller, 1999) or the crossing problem.

To help address these shortcomings, we are using the Principal Analysis by Conditional Expectation (PACE) (Yao et al., 2005), a functional data analysis approach, to model trajectories of proportional gray matter (pGM), white matter (pWM) and cerebrospinal fluid (pCSF) volumes from data from 446 children from 1 to 120 months of age. This method is specifically geared towards sparse and irregularly observed longitudinal data (Wang et al., 2016; Yao et al., 2005) and can be utilized to obtain estimates of true trajectories at subject level, which is a difficult task when most of the subjects have only few repeats. Combining these data and the recently developed Fréchet regression approach (Petersen and Müller, 2019) to construct age dynamic growth percentiles at the population level, we develop normative percentile 'brain growth charts' for proportions of GM, WM and CSF volumes from infancy to pre-adolesence. We then examine dynamic associations between brain growth trajectories and cognitive scores derived from the Mullen Scales of Early Learning (Mullen, 1995) in early and late childhood.

2 Materials and Methods

2.1 Subject Details and Demographics

Data used in this study were drawn from the ongoing longitudinal RESONANCE study of healthy and neurotypical brain and cognitive development, based at Brown University in Providence, RI, USA. From the RESONANCE cohort, 446 typically-developing children (195

girls) ages one to one hundred and twenty months were selected for analysis in this study. General participant demographics are provided in Table 1, with children being representative of the RI population.

RESONANCE is an accelerated-longitudinal study of a large community cohort of healthy children with approximately half of the cohort enrolled between two and eight months of age; and the remainder between two and four years of age. Depending on child age, study visits occur every six (under age two) or twelve months (over age two), and include multi-modal MRI, performance and parent-reported measures of cognitive and behavioral functioning, anthropometry, and biospecimen collection.

To focus on healthy and neurotypically developing children, those with known major risk factors for developmental abnormalities or cognitive impairments were excluded at enrollment. Specifically, children born preterm (< 37 weeks) or small for gestation age (< 1500 g), *in utero* exposure to alcohol, cigarette smoke or illicit substances; fetal ultrasound abnormalities; complicated delivery resulting in 5 minute APGAR scores < 8 and/or NCU admission; neurological disorder in child (e.g. head injury resulting in loss of consciousness, epilepsy); and psychiatric or learning disorder in the infant, parents or siblings (including maternal depression requiring medication in the year prior to pregnancy). In addition to screening at the time of enrollment, on-going screening for worrisome behaviors using validated tools were performed to identify at-risk children and remove them from subsequent analysis.

2.2 Ethics Statement

Research ethics oversight was provided by the host institutions, including the Brown University and Lifespan institutional review boards. For all children, written informed consent was obtained from their parents or legal guardians.

2.3 MRI Acquisition & Analysis

For all MR acquisition, children under 4 years of age were scanned during natural and non-sedated sleep and older children were imaged whilst watching a movie or other video. Our imaging protocol included relaxometry, multi-shell diffusion, resting-state connectivity, and

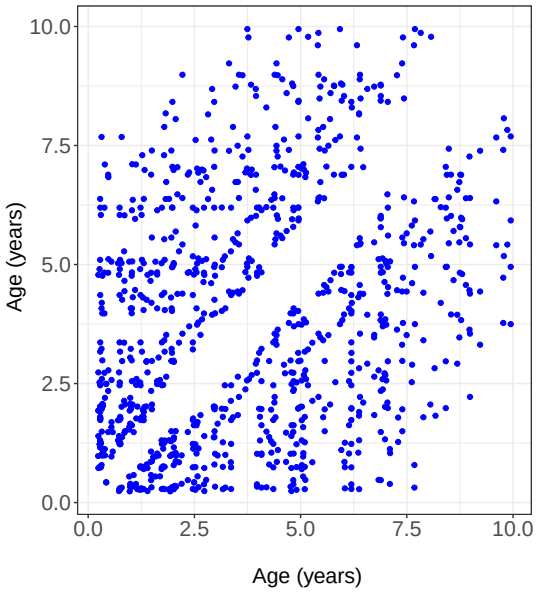

Figure 1: Design plot displaying all pairs of ages for all included children from the RESONANCE cohort.

magnetic resonance spectroscopy acquisitions in addition to the anatomical data. As a result, depending on child compliance (sleeping and/or motion) high quality anatomical data was not collected or available for every child at every scan time-point. Following data acquisition, all scans were inspected for motion-related artifacts and image blurring or ghosting.

The design plot in Figure 1 illustrates all pairwise measurements with age. For the RESONANCE cohort, this plot reveals the sparsity of the times when measurements were taken. Most of the children only had one scan (Table 1), and most of the measurements were taken at early ages before 5 years old (Figure 2).

Table 1: Distribution of numbers of repeats per child.

Number of repeats per child	1	2	3	4	5	6	8
Girls	114	44	21	12	4	0	0
Boys	149	54	24	14	6	4	1

T1-weighted anatomical data were acquired on a 3T Siemens Trio scanner with a 12-channel head RF array. T1-weighted magnetization-prepared rapid acquisition gradient echo anatomical data were acquired with an isotropic voxel volume of $1.2\times1.2\times1.2\text{mm}^3$, resampled to $0.9\times0.9\times0.9\text{mm}^3$ Sequence specific parameters were: TE = 6.9ms; TR = 16ms; inversion

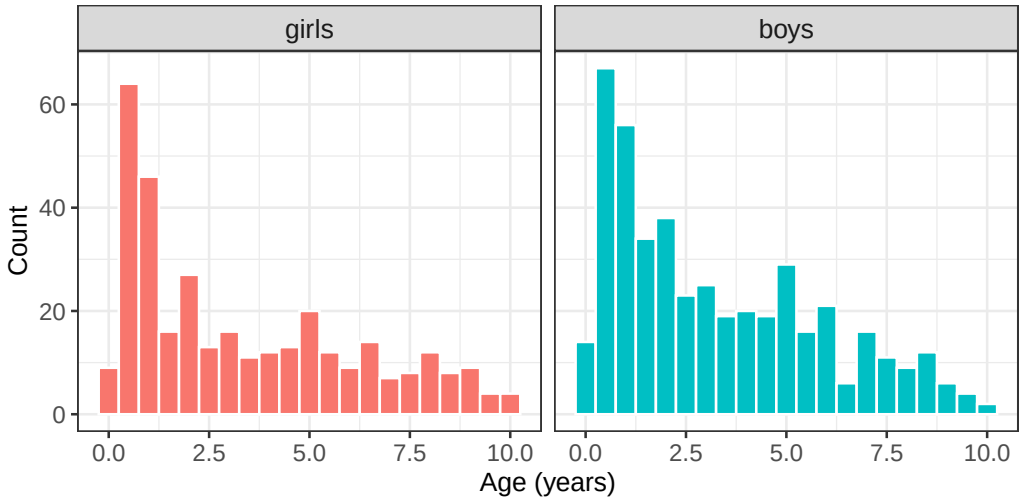

Figure 2: Distributions of ages of observations for girls (left) and boys (right).

preparation time = 950ms; flip angle = 15 degrees; BW = 450Hz/Pixel. The acquisition matrix and field of view were varied according to child head size in order to maintain a constant voxel volume and spatial resolution across all ages (Dean et al., 2014). Using a multistep registration procedure (O'Muircheartaigh et al., 2014), a series of study- and age-specific anatomical T1-weighted templates were created corresponding to 6, 9, 12, 15, 18, 21, 24, and 27 month ages. At least 10 boys and 10 girls were included in each template. An overall study template was then created from these age templates, which was aligned to the MNI152 template (Lancaster et al., 2007). Each child's anatomical T1-weighted image was transformed into MNI space by first aligning to their age-appropriate template and then applying the pre-computed transformation to MNI space, with the calculated individual forward and reverse transformations saved and used for the volumetric analysis described below. All template creation and image alignment was performed using a 3D nonlinear approach (ANTS, Avants et al., 2014) with cross-correlation and mutual information cost functions. This step was done so that previously calculated brain masks and initial WM, GM, and CSF estimates could be aligned to the each child's individual anatomical data and used as starting priors for the Atropos voxel-wise WM, GM, and CSF segmentation method. The resultant tissue partial volume maps were then thresholded at 0.2 and summed to calculate total-brain, WM, GM, and CSF volumes and their proportional fraction (i.e., pWM = WM/(WM+GM+CSF)). Overall, pGM and pCSF decrease and pWM increase as

children age (Figure 3).

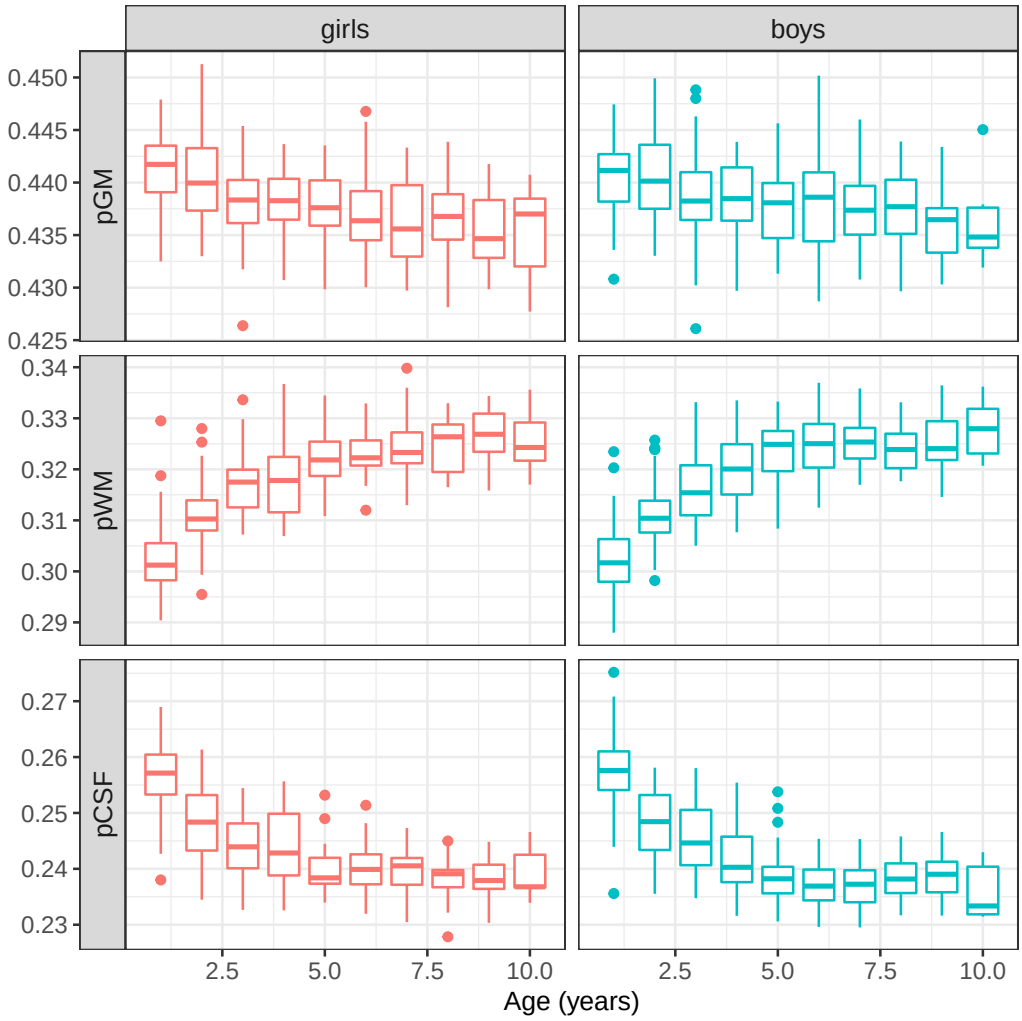

Figure 3: Distributions of pGM (top), pWM (middle), and pCSF (bottom) observed within each year increment of ages for girls (left) and boys (right).

2.4 Neurocognitive Assessments

Alongside neuroimaging data, each child's cognitive development was assessed using a combination of observed performance and parent-reported measures. For overall cognitive functioning, children under 5 years of age were assessed using the Mullen Scales of Early Learning (Mullen, 1995), a standardized and population-normed tool for assessing overall (Early Learning Composite, ELC), verbal (Verbal Development Quotient, VDQ) and non-verbal abilities (Non-Verbal Development Quotient, NVDQ). To assess overall cognitive function-

ing in older children, we used the full scale IQ (FSIQ) from the Wechsler Intelligence Scale for Children 4th edition (Wechsler, 2012), an individually administered standard intelligence test for children aged 6 to 16 years.

2.5 Statistical Methods

Functional data analysis (FDA) provides a powerful toolkit for analyzing longitudinal data. The idea is to view the measurements for each individual as values of a random trajectory, sometimes contaminated with measurement error. Similar to traditional principal component analysis, functional principal component analysis (FPCA) is typically used for dimension reduction and to identify dominant modes of variation in functional data. Classical FPCA (Wang et al., 2016; Hall et al., 2006) aims at fully observed or densely observed curves but faces great challenges in longitudinal settings when one has only very few repeated measurements for each subject or the measurements are on an irregular grid. While the mean function can be estimated by smoothing across neighborhoods even in sparse settings as depicted in Figure 1, the estimation of the covariance surface, which is the backbone of FPCA, is more complex in sparse settings (Yao et al., 2005; Wang et al., 2016). While this challenge has been addressed and forms the key to link functional and longitudinal data analysis, these developments are not widely known outside of nonparametric statistics and one of the main goals of this paper is to introduce this novel approach to researchers in child development, where longitudinal studies are paramount.

Principal component analysis through conditional expectation (PACE) (Yao et al., 2005) is geared towards situations where the study design is very sparse and irregular, which is often typical for longitudinal neuroimaging studies and is also the case for the RESONANCE data. By pooling observations across subjects followed by smoothing steps, one can get estimates of mean and covariance functions at the population level. At the subject level, one may then use the functional principal component scores obtained from the PACE approach to reconstruct trajectories. In Section 2.5.1, we describe in detail how we obtained mean and covariance functions, the eigenfunctions of the corresponding auto-covariance operator, the corresponding modes of variation and the individual trajectory fitting for the pGM, pWM and pCSF volumes acquired in RESONANCE.

The software to implement PACE modeling is available on CRAN as an R package titled `fdapace` at https://CRAN.R-project.org/package=fdapace; see Carroll et al. (2020). We use the functional concurrent regression model, also known as the varying coefficient model, as described in Section 2.5.2, to study the association of pGM, pWM and pCSF with the cognitive scores. Various estimation techniques are available for the model parameters such as slope (function) and the intercepts for both dense and sparse longitudinal data (Cai et al., 2000; Huang et al., 2004; Şentürk and Müller, 2010; Wu et al., 2010; Dai et al., 2019).

To derive and construct conventional quantile growth charts for pGM, pWM and pCSF volumes, we employ local Fréchet regression (Petersen and Müller, 2019) which is a non-parametric regression method with responses (in our case probability distributions) lying in metric spaces and Euclidean predictors, whence we obtain estimates for age-dependent quantile functions and hence age-varying dynamic quantiles/percentiles. An R package `frechet` for the implementation of local Fréchet regression is available on CRAN at https://CRAN.R-project.org/package=frechet (Chen et al., 2020).

2.5.1 PACE Modeling

FPCA is a dimension reduction method that summarizes functional data in the form of scalar valued functional principal component (FPC) scores. For a random function $X(t), t \in \mathcal{I}$, the mean function $\mu(t)$ and the autocovariance surface $C(s,t)$ are given by

$$\mu(t) = \mathbb{E}\left(X(t)\right) \text{ and } C(s,t) = \text{Cov}\left(X(s), X(t)\right) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s)\phi_k(t),$$

where $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant 0$ are the eigenvalues and ϕ_k are the orthonormal eigenfunctions of the autocovariance operator given by $\mathcal{C} : L_2(\mathcal{I}) \to L_2(\mathcal{I})$, $\mathcal{C}(f) = \int_{\mathcal{I}} C(s,t)f(s)ds$. By the Karhunen–Loève expansion, one can represent $X(t)$ as

$$X(t) = \mu(t) + \sum_{k=1}^{\infty} \psi_k \phi_k(t), \ t \in \mathcal{I},$$

where $\psi_k = \int_{\mathcal{I}}(X(t) - \mu(t))\phi_k(t)dt$ are the functional principal components. The ψ_k are zero mean uncorrelated random variables, accounting for random fluctuations of the trajec-

tory $X(t)$ around the mean curve $\mu(t)$. For the purpose of dimension reduction, the first K eigenfunctions are used so that $X(t) \approx \mu(t) + \sum_{k=1}^{K} \psi_k \phi_k(t)$ is represented using FPCs $(\psi_1, \ldots, \psi_K)$.

In practice, one has to estimate $\mu(t)$, $C(s,t)$, λ_k and ϕ_k using a sample of observations $\tilde{X}_i(t_j)$ where

$$\tilde{X}_i(t_j) = X_i(t_{ij}) + \epsilon_{ij}$$

are noisy longitudinal data, a corrupted version of $X_i(t)$ observed on a grid $t_{i1}, t_{i2}, \ldots, t_{in_i}$ for the i^{th} subject. The noise variables ϵ_{ij} are assumed to be normal with zero mean and variance σ^2. When the grid is the same and the n_i are large for all the subjects, the cross-sectional estimates of the targets $\hat{\mu}(t)$, $\hat{C}(s,t)$, $\hat{\lambda}_k$, $\hat{\phi}_k$, $\hat{\sigma}^2$ are reliable, leading to trustworthy estimates for ψ_k. However if the n_i are small or the grid is irregular, as is typical for many longitudinal studies in child development, estimation is much more challenging and the numerical integration step in the estimation of the FPC scores ψ_k does not work any longer. In such situations, the mean curve estimate $\hat{\mu}(t)$ is obtained by smoothing pooled data from all subjects and the covariance function estimate $\hat{C}(s,t)$ is obtained by smoothing raw covariances (Yao et al., 2005). For the RESONANCE data, to adjust for their extreme sparsity, we use 10% of the age domain and 15% of the age domain as bandwidth choices for mean and covariance smoothing, respectively. Finally the FPCs ψ_k are obtained using best linear unbiased predictors

$$\hat{\psi}_k = \hat{\lambda}_k \hat{\phi}_k^T \hat{\Sigma}_{X_i}^{-1} (\tilde{X}_i - \hat{\mu}) \tag{2}$$

where

$$\tilde{X}_i = (X_i(t_{i1}), \ldots, X_i(t_{in_i}))\,,\ \hat{\phi}_k = \left(\hat{\phi}(t_{i1}), \ldots, \hat{\phi}(t_{in_i})\right),\ \hat{\mu} = (\hat{\mu}(t_{i1}), \ldots, \hat{\mu}(t_{in_i}))$$

and

$$\{\hat{\Sigma}_{X_i}\}_{uv} = \hat{C}(t_{iu}, t_{iv}) + \hat{\sigma}^2 I_{n_i}.$$

To determine the number of components K, it is common to take the minimum number for explaining a given fraction of variance explained by looking at $\sum_{k=1}^{K} \hat{\lambda}_k / \sum_{k=1}^{\infty} \hat{\lambda}_k$. Once

K has been chosen, individual trajectories are reconstructed as

$$X_i(t) = \hat{\mu}(t) + \sum_{k=1}^{K} \hat{\psi}_k \hat{\phi}_k(t).$$

Under additional Gaussian assumptions, the estimates of the predicted FPC scores converge to their targets, and pointwise/uniform confidence bands can be constructed for predicted trajectories (Yao et al., 2005). In practical applications, to obtain the inverse in (2) is sometimes problematic; a ridge approach that provides a generalized inverse can overcome this and is described in Section S.1 in the Supplementary Materials.

For visualization in FDA, the modes of variation provide an insightful representation of variance decomposition in a sample of functional or longitudinal data. The modes of variation capture the deviation around the mean function scaled by the shape of the dominant eigenfunctions. Formally, the k^{th} mode of variation is defined as

$$\mu(t) \pm k\sqrt{\lambda_k}\phi_k(t),$$

where the population quantities are estimated using their sample counterparts in data applications.

2.5.2 Varying Coefficient Modeling

The varying coefficient regression model, or the functional concurrent regression model, between functional response $Y(t)$ and functional predictor $X(t)$ is given by

$$Y(t) = \beta_0(t) + \beta(t)X(t) + \epsilon(t) \tag{3}$$

where $\beta_0(t)$ and $\beta(t)$ are smooth coefficient functions and $\epsilon(\cdot)$ is a zero mean Gaussian process. Various estimation techniques are available for the intercept function $\beta_0(t)$ and the slope function $\beta(t)$ for both dense and sparse functional data (Cai et al., 2000; Huang et al., 2004; Şentürk and Müller, 2008; Wu et al., 2010; Şentürk and Müller, 2010; Şentürk and Nguyen, 2011; Huang et al., 2004). For assessing the strength of the association, one can use

the time-varying R^2 function

$$R^2(t) = 1 - \frac{\mathrm{Var}(\epsilon(t))}{\mathrm{Var}(Y(t))}.$$

Greater values of $R^2(t)$ indicate that more variability in the response curve $Y(t)$ is explained by the linear model in $X(t)$. Positive values of the slope function $\beta(t)$ indicate positive association between $Y(t)$ and $X(t)$ at t while negative slopes imply negative association.

2.5.3 Dynamic Quantile Modeling using Local Fréchet Regression

Given a closed interval $D \subset \mathbb{R}$, we focus on the Wasserstein space $\mathcal{W} = \mathcal{W}(D)$ of probability distributions on D with finite second moments, endowed with the $\mathcal{L}^2$ Wasserstein distance

$$d_W(q_1, q_2) = \left\{ \int_0^1 [F_1^{-1}(u) - F_2^{-1}(u)]^2 \mathrm{d}u \right\}^{1/2}, \quad \text{for } q_1, q_2 \in \mathcal{W}.$$

Here, F_l and F_l^{-1} are the cumulative distribution function (cdf) and quantile function of q_l, for $l = 1, 2$, where quantile functions are considered to be the left continuous inverse of the corresponding cdfs; specifically, given a cdf F,

$$F^{-1}(u) = \inf\{x \in D : \ F(x) \geqslant u\}, \quad \text{for } u \in (0, 1).$$

Let (T, P) be a pair of random elements taking values in $\mathcal{T} \times \mathcal{W}$ with joint distribution $\mathcal{F}$, where $\mathcal{T} \subseteq \mathbb{R}$ is the time domain. Due to the compactness of D, $\mathbb{E}[d_W^2(P, q) \mid T = t] < \infty$, for all $q \in \mathcal{W}$ and $t \in \mathcal{T}$. Since the Wasserstein space $\mathcal{W}$ is a Hadamard space (Kloeckner, 2010), there exists a unique minimizer of $\mathbb{E}[d_W^2(P, \cdot) \mid T = t] =: M(\cdot, t)$ (Sturm, 2003), which is the conditional Fréchet mean $\mu_\oplus(t)$ of P given $T = t$. Specifically,

$$\mu_\oplus(t) = \underset{q \in \mathcal{W}}{\operatorname{argmin}}\, M(q, t).$$

With $\mu_\oplus(t)$, the τth conditional quantiles of P given $T = t$ can be expressed as $F^{-1}_{\mu_\oplus(t)}(\tau)$, for $\tau \in (0, 1)$, where $F^{-1}_{\mu_\oplus(t)}$ is the quantile function of $\mu_\oplus(t)$. Thus, in order to estimate the conditional quantiles, it suffices to estimate the conditional Fréchet mean $\mu_\oplus(t)$. To this end, we employ the local Fréchet regression approach (Petersen and Müller, 2019), described as

follows. Firstly, we define a localized Fréchet mean by

$$\nu_{\oplus}(t) = \underset{q \in \mathcal{W}}{\operatorname{argmin}} L_n(q,t), \quad \text{with } L_n(q,t) = \mathbb{E}[w(T,t,h)d_W^2(P,q)].$$

Here, $w(s,t,h) = \sigma_0^{-2} K_h(s-t)[\kappa_2 - \kappa_1(s-t)]$, where $\kappa_z = \mathbb{E}[K_h(T-t)(T-t)^z]$, for $z = 0,1,2$, $\sigma_0^2 = \kappa_0\kappa_2 - \kappa_1^2$, $K_h(\cdot) = K(\cdot/h)/h$, K is a smoothing kernel, i.e., a density function symmetric around zero, and $h = h(n) > 0$ is a bandwidth sequence.

Suppose $\{(T_i, P_i)\}_{i=1}^n$ are independent realizations of (T,P). If the distributions P_i are fully observed, setting $\widehat{w}(s,t,h) = \widehat{\sigma}_0^{-2} K_h(s-t)[\widehat{\kappa}_2 - \widehat{\kappa}_1(s-t)]$, where $\widehat{\kappa}_z = n^{-1}\sum_{i=1}^n K_h(T_i - t)(T_i - t)^z$, for $z = 0,1,2$, and $\widehat{\sigma}_0^2 = \widehat{\kappa}_0\widehat{\kappa}_2 - \widehat{\kappa}_1^2$, an oracle local Fréchet regression estimate is

$$\widetilde{\nu}_{\oplus}(t) = \underset{q \in \mathcal{W}}{\operatorname{argmin}} \widetilde{L}_n(q,t), \quad \text{with } \widetilde{L}_n(q,t) = n^{-1}\sum_{i=1}^n \widehat{w}(T_i,t,h)d_W^2(P_i,q).$$

In practice, the distributions P_i are however rarely fully observed; instead we only observe random samples of measurements generated from P_i. This issue can be addressed by estimating cdfs (e.g., Aggarwal, 1955; Read, 1972; Falk, 1983; Leblanc, 2012), quantile functions (e.g., Parzen, 1979; Falk, 1984; Yang, 1985; Cheng and Parzen, 1997) or density functions (e.g., Panaretos and Zemel, 2016; Petersen and Müller, 2016) of the underlying distributions. For any $q \in \mathcal{W}$, we denote the estimated distribution by $\widehat{q} = q(\widehat{F})$, where $\widehat{F}$ is a cdf estimate based on a random sample generated from q. If employing quantile function or density estimation methods, one can obtain the cdf estimate by right continuous inversion or integration. Replacing P_i with the corresponding estimates $\widehat{P}_i$, a data-based local Fréchet regression estimate is

$$\widehat{\nu}_{\oplus}(t) = \underset{q \in \mathcal{W}}{\operatorname{argmin}} \widehat{L}_n(q,t), \quad \text{with } \widehat{L}_n(q,t) = n^{-1}\sum_{i=1}^n \widehat{w}(T_i,t,h)d_W^2(\widehat{P}_i,q).$$

However, in the RESONANCE data, only one measurement is available per distribution at most ages. Subsequently, the distribution estimation methods mentioned before cannot be applied directly. To resolve this issue, we divide the age domain $\mathcal{T}$ into bins $S_1 = [a_0, a_1), S_2 = [a_1, a_2), \ldots, S_B = [a_{B-1}, a_B]$, with $\min(\mathcal{T}) = a_0 < a_1 < \ldots < a_B = \max(\mathcal{T})$. Pooling the observations from all the subjects together, we denote the paired observations

by (t_j, Y_j), where t_j is the observed time and Y_j is a measurement generated from a random distribution $P(t_j)$ at age t_j, for $j = 1, \ldots, m$. For the mid age $b_k = (a_{k-1} + a_k)/2$ of each bin, we can obtain an estimate $\breve{P}(b_k)$ of the distribution $P(b_k)$ based on the observations $\{Y_j : t_j \in S_k\}$ obtained at ages falling within the bin. Then the local Fréchet regression estimate for RESONANCE data is

$$\breve{\nu}_{\oplus}(t) = \underset{q \in \mathcal{W}}{\operatorname{argmin}} \breve{L}_B(q, t), \quad \text{with } \breve{L}_B(q, t) = B^{-1} \sum_{k=1}^{B} \widehat{w}(b_k, t, h) d_W^2(\breve{P}(b_k), q). \tag{4}$$

With $\breve{\nu}_{\oplus}(t)$ for $t \in \mathcal{T}$, for any given $\tau \in (0, 1)$, an estimate for the τth conditional quantile curve g_τ is

$$\widehat{g}_\tau(t) = F^{-1}_{\breve{\nu}_{\oplus}(t)}(\tau), \quad \text{for } t \in \mathcal{T}, \tag{5}$$

where $F^{-1}_{\breve{\nu}_{\oplus}(t)}$ is the quantile function of $\breve{\nu}_{\oplus}(t)$.

For RESONANCE data, we set the number of bins $B = 40$, where the bins are constructed such that each bin contains about $1/B \times 100\% = 2.5\%$ of the observed data. The bandwidth h was chosen by 10-fold cross validation simultaneously for pGM, pWM and pCSF; specifically, the bandwidth used for the analysis of proportions is 1.65.

3 Results

3.1 Population Level Analysis

3.1.1 Mean Function

Figure 4 illustrates the population level mean curves for development of proportions pGM, pWM and pCSF in children from one year to nine years of age. We constructed 95% pointwise confidence intervals by resampling the subjects $5,000$ times to generate the bootstrap replicates of the mean function and then obtained the pointwise cutoffs from the bootstrap replicates. No major differences in the levels of the mean proportions were detected between boys and girls (Figure 4). However, there are significant differences in the mean raw brain volume levels between boys and girls as illustrated in Figure S.1 in the supplement.

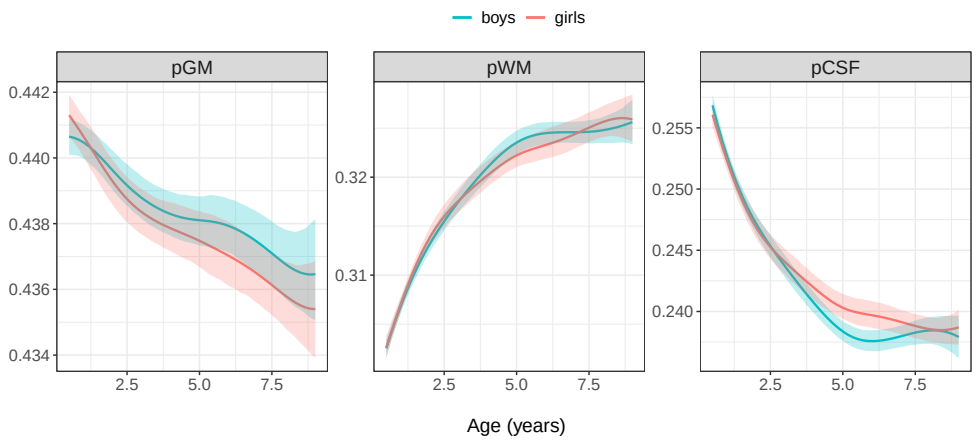

Figure 4: Population level mean functions (solid lines) for proportion of gray matter (left), white matter (middle) and CSF (right) evolution for children in the RESONANCE cohort. The light ribbons correspond to 95% pointwise confidence intervals.

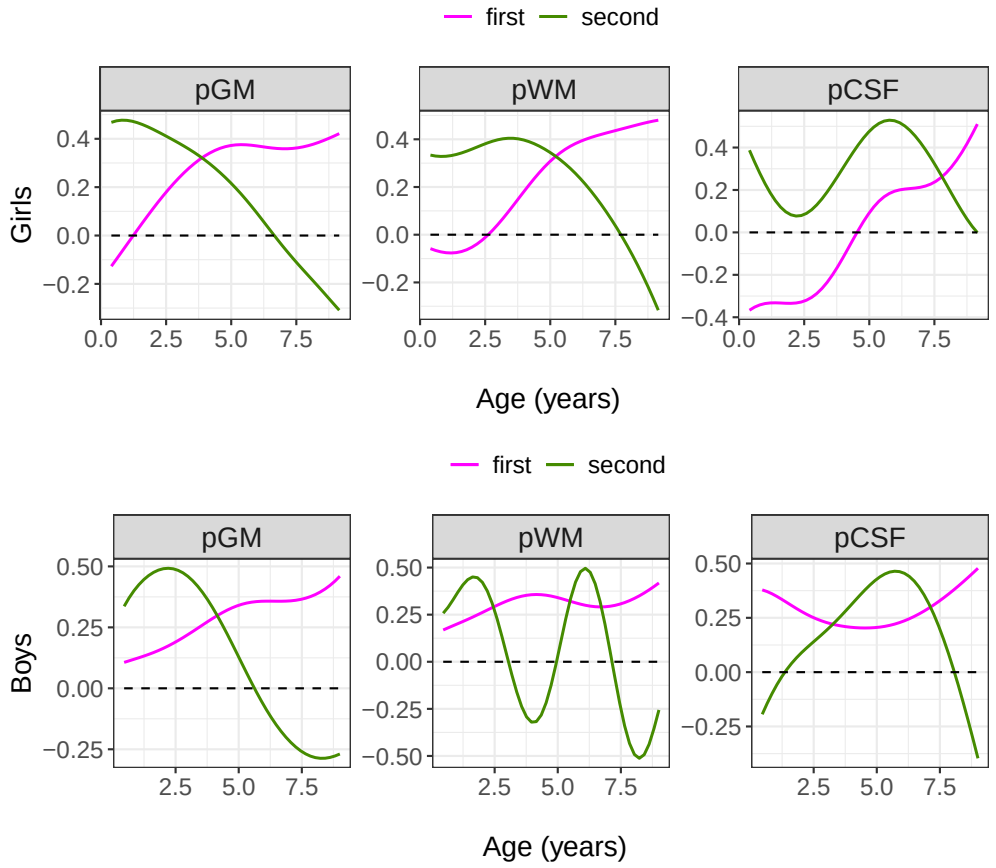

Figure 5: Eigenfunctions for FPCA of proportion of gray matter (left), white matter (middle) and CSF (right) for girls (top panel) and boys (bottom panel) in the RESONANCE cohort.

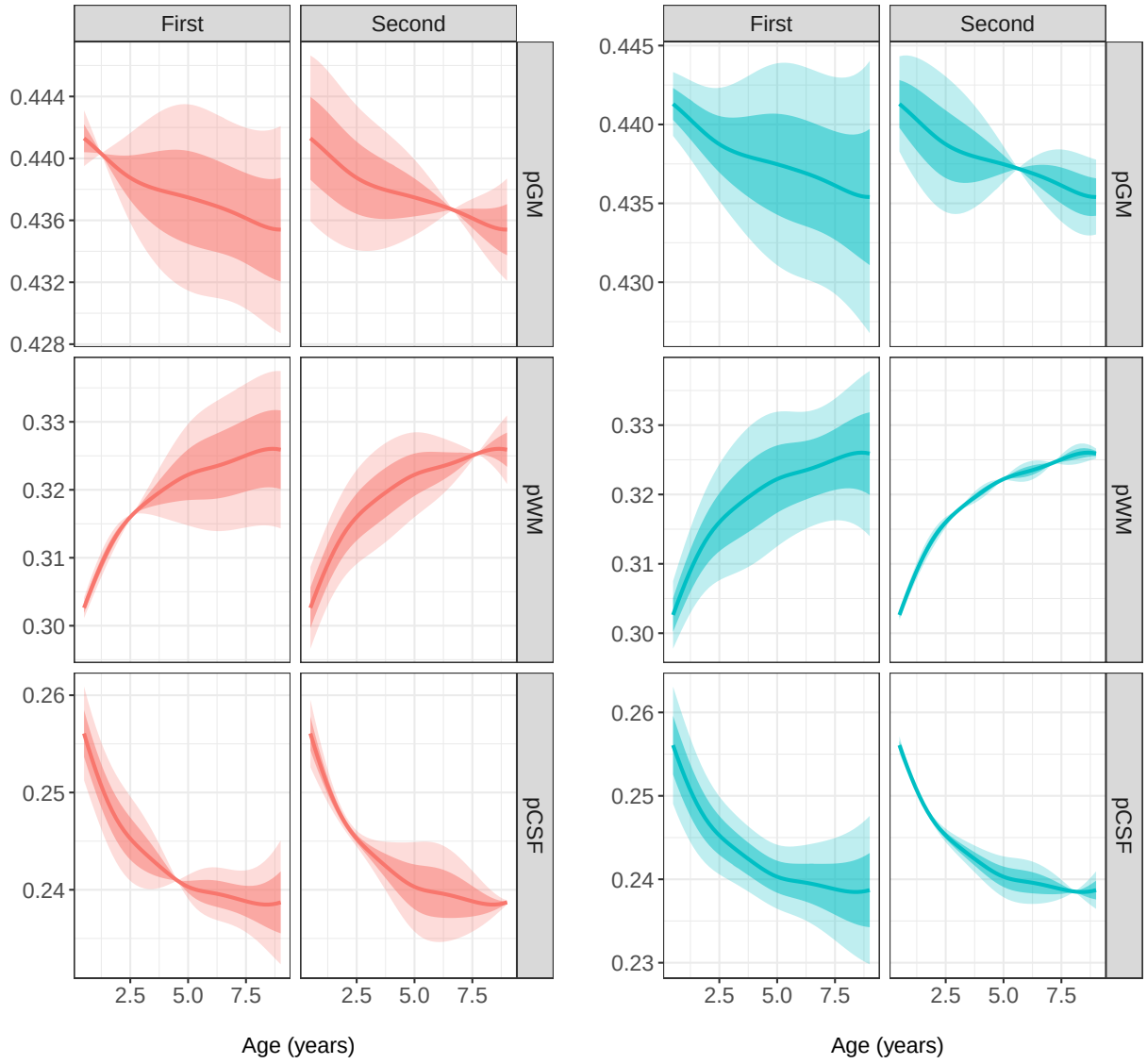

Figure 6: Mode of variation plot for proportion of gray matter (top), proportion of white matter (middle) and proportion of CSF (bottom) for girls (left) and boys (right) in the RESONANCE cohort. The solid line corresponds to the mean function, the darker ribbon represents the area covered by the first mode of variation around the mean function and the lighter ribbon represents the area covered by the second mode of variation around the mean function.

Figure 5 shows the dominant eigenfunctions and Figure 6 the corresponding modes of variation obtained using the PACE approach of covariance estimation for the children in the RESONANCE cohort differentiated by gender. For girls, the first eigenfunction for proportions of gray matter and white matter is non-negative and increasing for most of the period between one to nine years. This indicates that the predominant source of variation in

the trajectories of pGM and pWM arises from increasing departures of these measurements from the mean value as the children grow up. For pCSF in girls, the predominant variation is reflected in contrasting patterns of growth between early and later ages. These shapes capture the nature of variation in the trajectories around the mean function as reflected in the corresponding first mode of variation in Figure 6. The contrasting shapes between different growth periods forms the second mode of variation for pGM and pWM in girls and also for pCSF in boys. Figures S.3 and S.2 in the supplement show the corresponding behavior for the raw volumes.

3.1.2 Dynamic Percentiles

We constructed percentile growth charts of age-based dynamic quantiles by local Fréchet regression as per (5) for pGM, pWM and pCSF, as shown in Figure 7. The resulting dynamic percentiles for girls and boys mostly evolve in similar patterns. The dynamic percentiles of pGM and pCSF in general decrease while the curves for pWM increase as children age, in line with the crude summary of age-dependent distributions in Figure 3. Fluctuation in the dynamic percentile curves of pCSF for boys reduces dramatically around age 6, while such change seems to occur later for girls. For pWM, the augmentation of the dynamic percentiles is relatively fast before age 4 and subsequently slows down. The percentiles of pGM show similar temporal dynamics between genders overall, yet a difference can be seen in the 0.95 percentiles for higher ages. The results for volumes of GM, WM, and CSF are shown in Figure S.4 and total brain volumes (TBV) in Figure S.5 in the Supplementary Material. As a cautionary note we remark that the brain-for-age curves in this paper may not represent the brain development in early childhood of for the entire population of typically developing children in the US, or even in Providence, RI, due to the limited number of children involved in this study. Our main goal here is rather to provide and illustrate a useful method to construct brain-for-age curves for neurodevelopmental studies.

3.1.3 Dynamic Association with Cognitive Development scores

Brain structure and tissue development are known to be associated with cognitive skills. In the following, we fit a linear varying coefficient model as described in equation (3) of section

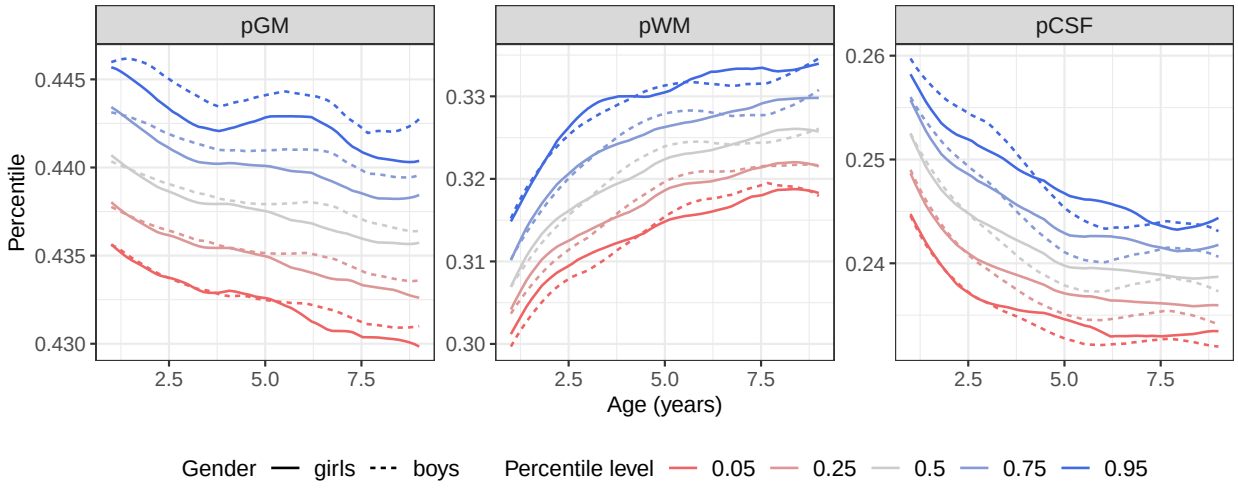

Figure 7: Age-based dynamic percentiles of proportions of GM (pGM, left), WM (pWM, middle) and CSF (pCSF, right) estimated by local Fréchet regression as per (5) for RESONANCE data.

2.5.2 with the time-varying ELC, VDQ, NVDQ and FSIQ scores as the response and pGM, pWM and pCSF as predictors for the age interval 2 to 5.5 years for ELC, VDQ and NVDQ scores and the age interval 6 to 10 years for the FSIQ score. Due to the extreme sparsity, we do not differentiate this analysis across genders.

Figure 8 illustrates that the early age cognitive score NVDQ is significantly negatively correlated with pCSF between three and a half to four and a half years of age. Around the same period, NVDQ also shows significant positive associations with pWM. ELC is also positively associated with pWM right before four years of age and negatively associated with pCSF starting from four years to slightly after four years. For later years, the FSIQ score is significantly negatively associated with pCSF around seven years of age and after nine years and positively associated with pWM after nine years.

3.2 Subject Level Analysis

3.2.1 Individual Trajectory Modeling

Using PACE, as described in section 2.5.1 of the supplement, one can reconstruct the individual smooth underlying trajectories at the subject level. Trajectory predictions can be obtained using the function `fitted` in the R package `fdapace` (Carroll et al., 2020). We

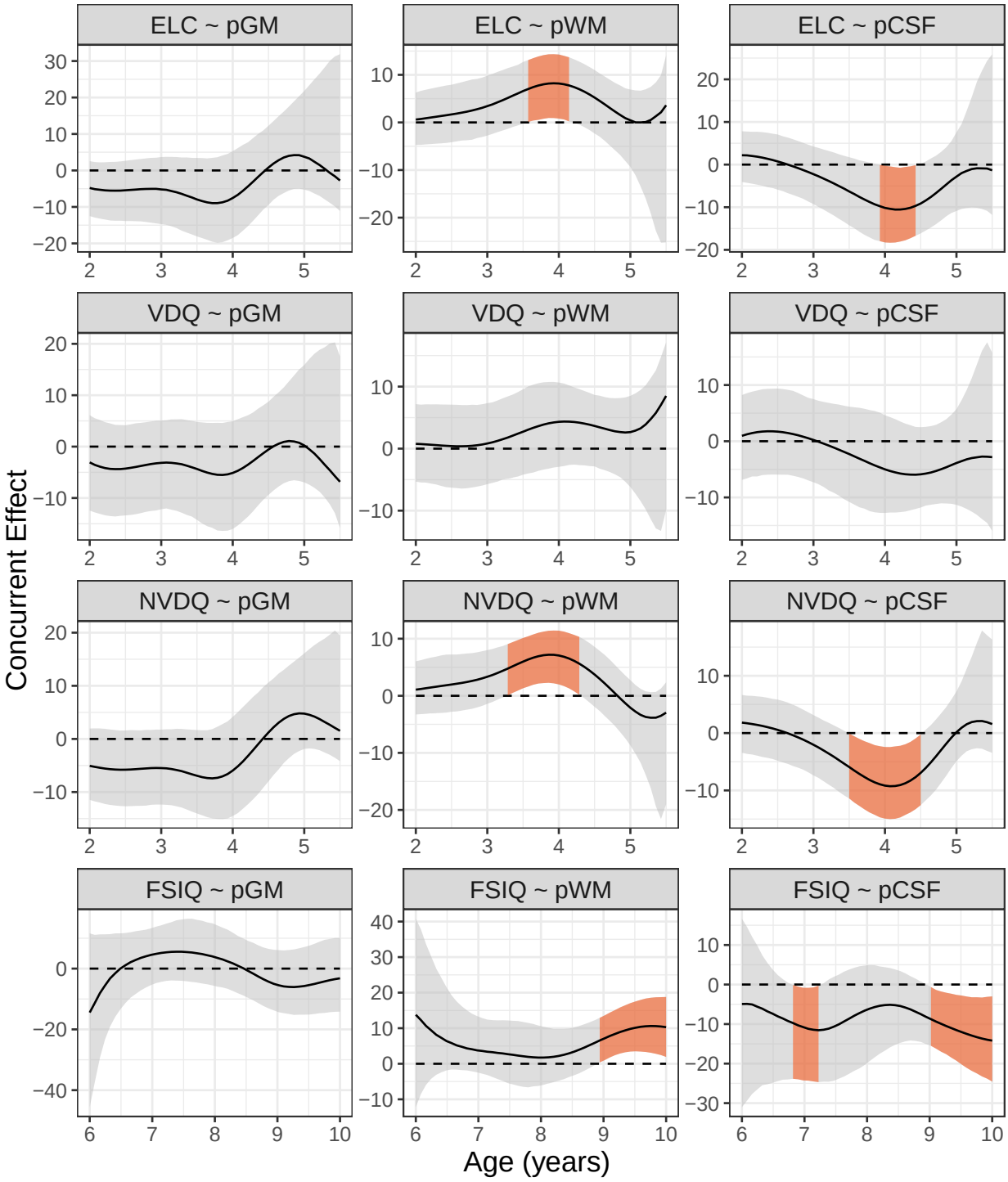

Figure 8: The regression slope function (black solid line) along with 95% pointwise confidence intervals (the grey ribbon) for varying coefficient model in equation (3) of section 2.5.2 with ELC, VDQ, NVDQ and FSIQ as response (top to bottom) and pGM (left), pWM (middle) and pCSF (right) as predictors. The orange portions of the ribbon indicate intervals where the slope function is significantly away from zero.

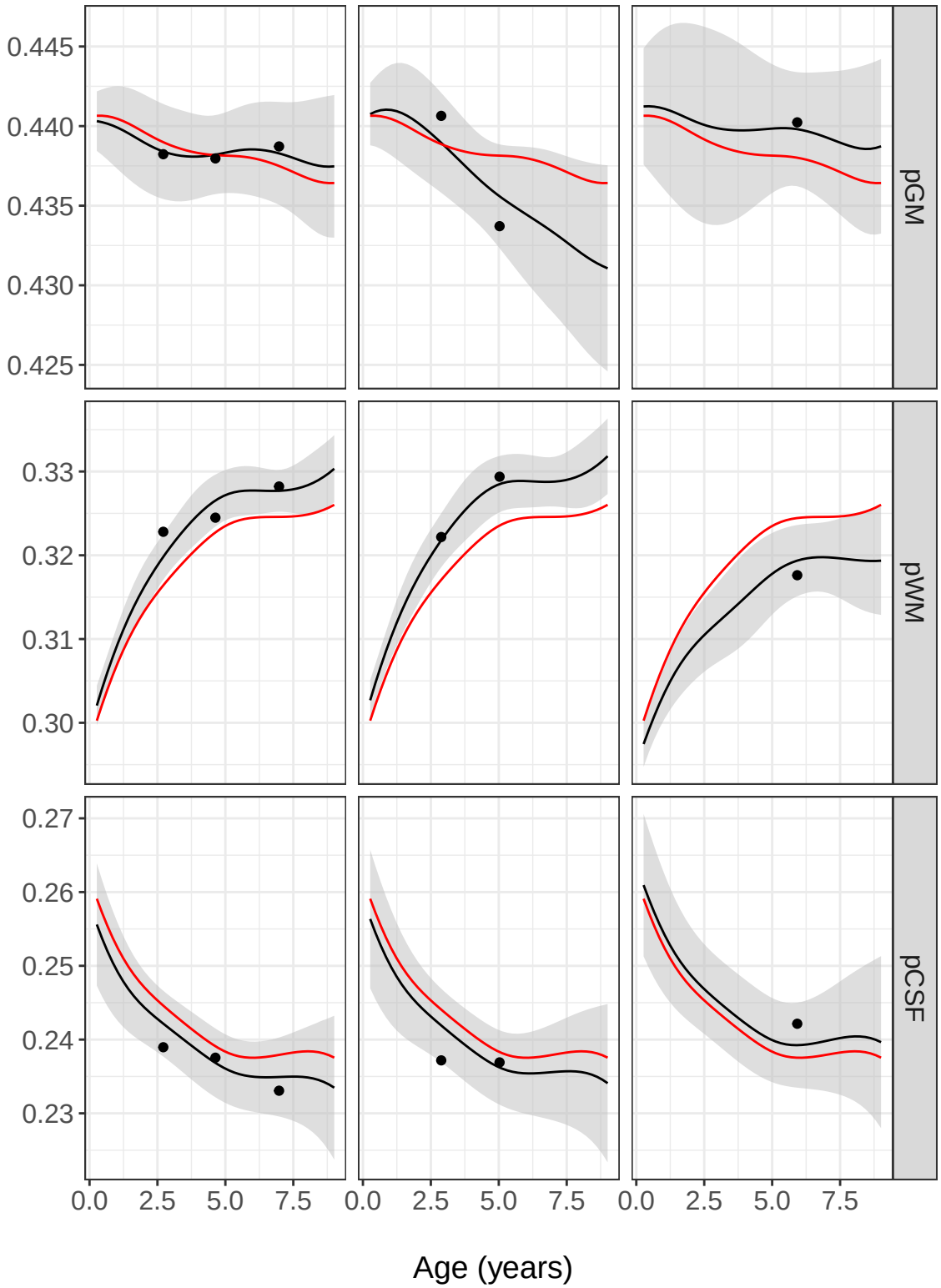

Figure 9: Individual trajectory prediction for three random boys in the RESONANCE data. The black solid curves correspond to the fitted trajectory and grey ribbon to the 95% simultaneous confidence band around it. The red curve represents the population mean curve and solid points correspond to the observations that were used in the fitting step.

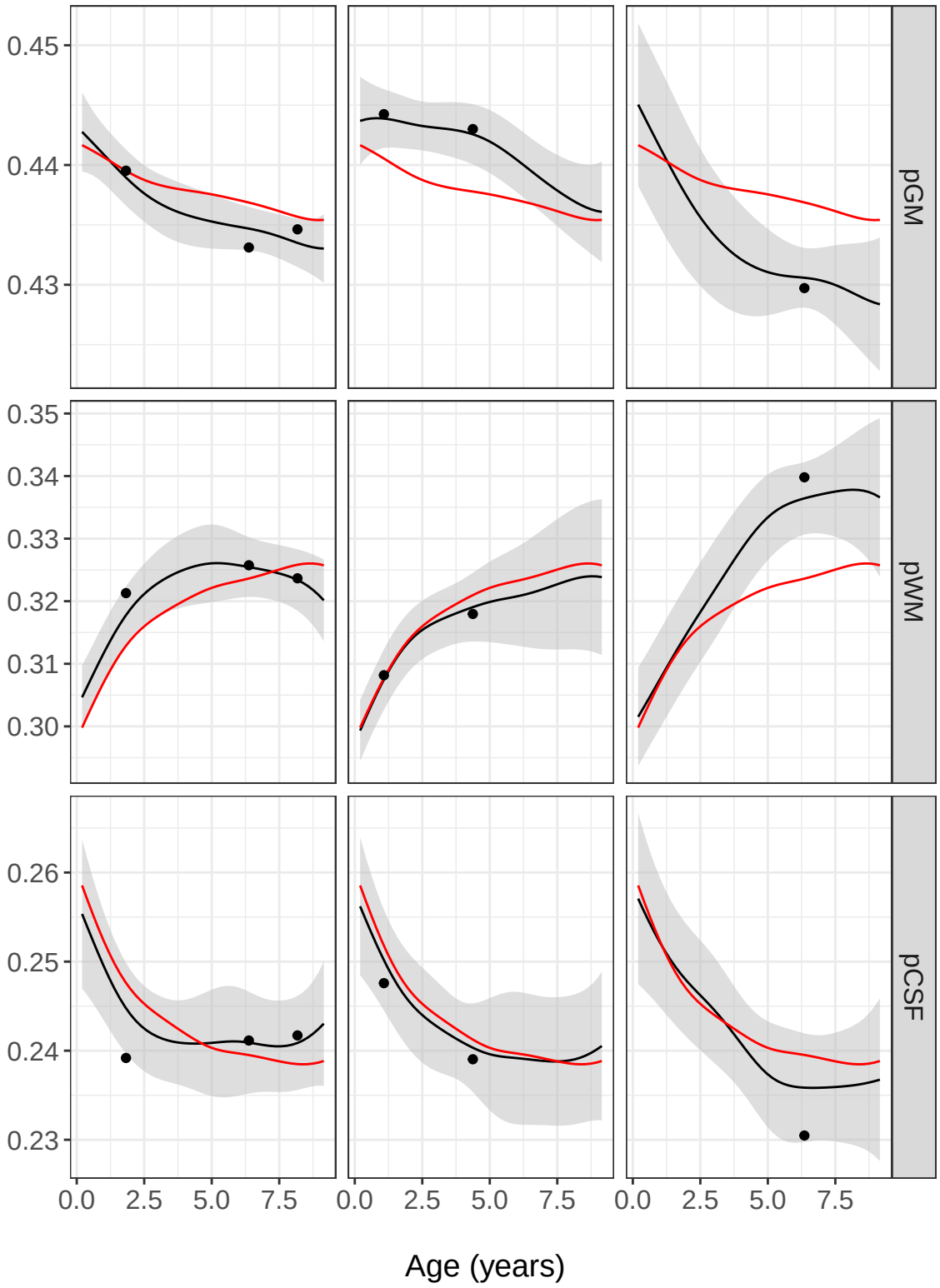

Figure 10: Individual trajectory prediction for three random girls in the RESONANCE data. The plotting convention is same as in Figure 9.

illustrate this approach for predicting the longitudinal evolution of pGM, pWM and pCSF volumes for three randomly selected boys in Figures 9 and three randomly selected girls in Figure 10. The plots show that the fitted trajectories align well with the measurements. The corresponding fits for the raw volumes of GM, WM, CSF, and TBV are illustrated in Figures S.6 and S.7 in the Supplement.

For validation of our approach, we selected one male and one female child with high scan frequency (5 and 8 respectively) and only used the first 50% of their measurements for fitting the trajectories. Once the fitted trajectories are obtained, the remaining measurements are compared to the predicted ones. Figure 11 demonstrates the results. The future measurements for these subjects are mostly contained within the uniform confidence band of the fitted trajectories, constructed under the Gaussian assumption, in most of the cases. This shows that the proposed approach does quite well in predicting individual trajectories even with sparse observations per child. The corresponding validation results on the absolute values of GM, WM, CSF and total brain volumes are illustrated in Figure S.8 in the Supplement.

4 Discussion

4.1 Population-based longitudinal brain development with age

One of the most commonly used models to define paediatric well-being and development are growth charts. Unfortunately, they only investigate outer physical features such as length and weight for age, features mediated in their association with cognitive functioning by brain size (Vuoksimaa et al., 2018). Healthy brain development has been identified as a key predictor of current and future cognitive development (for a review, Gilmore et al., 2018), but population-based childhood developmental brain-for-age growth charts are still missing. One of the reasons could be that contrary to outer physical features, which are easier to record, paediatric longitudinal MRI data are difficult to obtain and hence most often are observed sparsely in time. These also happens in scenarios where scan data are missing or obtained at a later time point (i.e. there is no concurrent acquisition), which then may lead

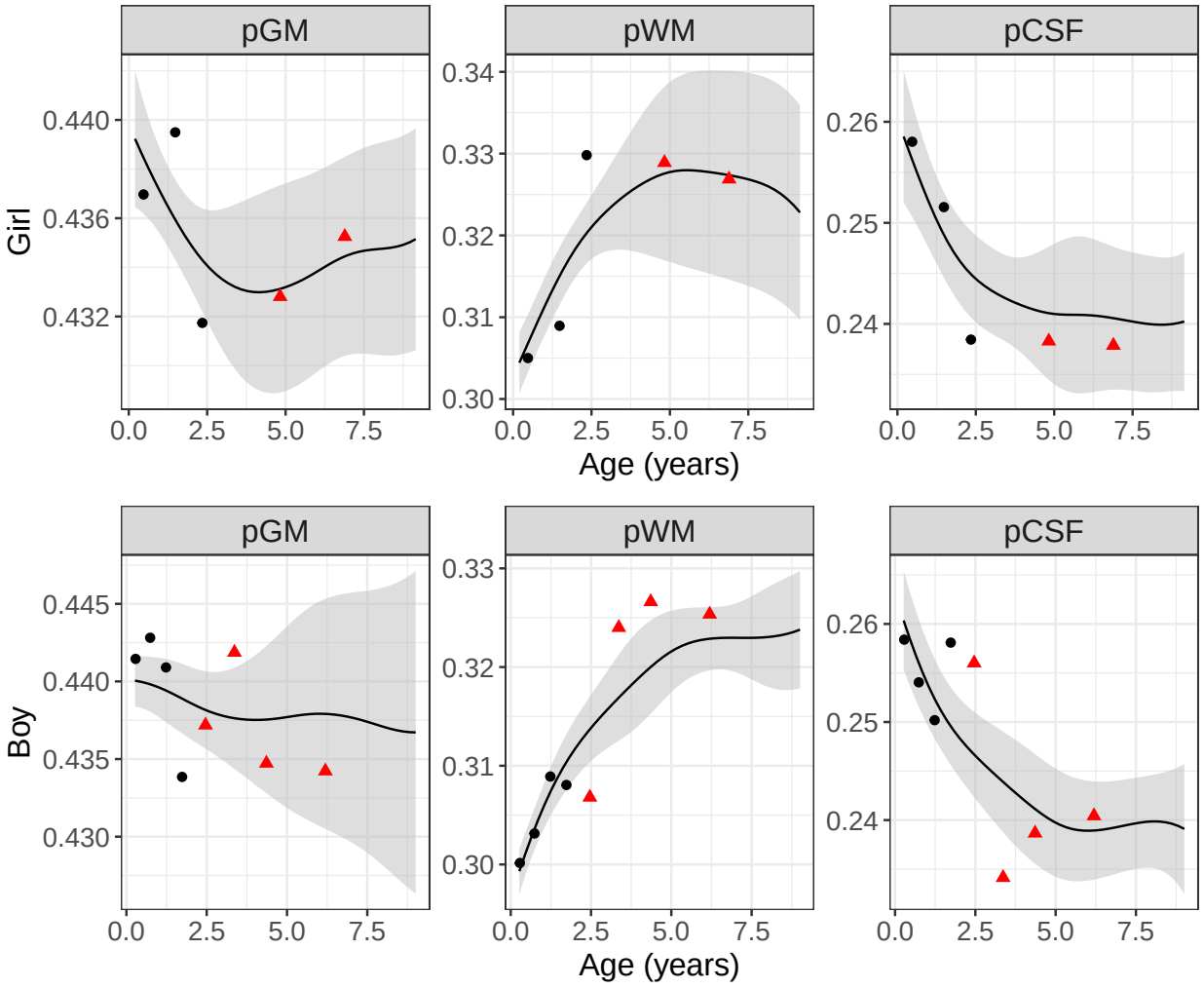

Figure 11: The black curves correspond to the fitted trajectory and the grey ribbon to the 95% simultaneous confidence band for one selected boy (top) and one selected girl (bottom). The black round points were used in the fitting step and the red triangular points are the future measurements of the same child.

to a participant's exclusion from the analysis or a shift to a cross-sectional approach.

In order to compare longitudinal modeling outputs from the most prominent modeling approaches, we first investigated average brain development with age at 95% point-wise confidence intervals (Figure 4), eigenfunctions (Figure 5), and modes of variation (Figure 6) estimated using PACE method; all divided by biological sex. As expected, raw volume measures increased in total brain volumes (TBV) as well as GM, WM and CSF with age (Figures S.1–S.3). Investigating proportional brain volumes, we demonstrated an initial pGM and pCSF decrease coupled with and pWM increase. The original pGM decline coupled with pWM increase reflects previous MR findings (Giedd et al., 1999; Toga et al., 2006; Brain Development Cooperative Group, 2012) and parallels cellular processes of pruning following prenatal neurogenesis and neural migration as expressed in pGM decline (and reflected in their distribution (Figure 9), and synaptogenesis in the brain (for a review, Silbereis et al., 2016). While trends were similar across methods, biological sex differences in proportional brain volume development with age were consistent too. Specifically, differences in later pCSF development seem to diverge between girls and boys, with stonger decreases that flatten around age 6 for boys, and less pronounced decreases with a later flattening at 7.5 years for girls.

However, while population based mean function models help describe overall development, they are unable to detect whether a child is well developing or struggling for their age. Like physical growth charts, population-derived brain percentile growth curves allow investigations of brain development across differing geographies and environmental settings. Here, we used Fréchet regression to create percentile growth charts of age-based dynamic quantiles for pGM, pWM and pCSF (Figure 7). The resulting dynamic percentiles for boys and girls mostly evolve in similar patterns (Figures 9–10). Differences within development that were not detectable previously are now more apparent. For example, the earlier flattening of the pCSF curve in boys appears to happen earlier in the lower percentiles than the upper percentiles (75th and above), and children in the 95th percentile experience a steeper second increase in pGM between ages 3.5 and 6.5 when compared to children in lower percentiles (Figure 7). Thus, using percentiles helps to place individual brain development trajectories, informing about "where on the curve" a child is relative to the population. This aids to flag

outliers and is expected to be useful for monitoring early brain development of individual children. For example, early brain overgrowth is a characteristic finding in autism disorder; and abnormalities in ventricle size may be indicative of inflammatory and other neurological disorders. Further, mapping of early brain developmental growth curves could be useful for identifying sensitive windows of changing dynamics that, for example, occur alongside major developmental milestones or the acquisition of new functional skills and abilities (e.g., crawling, walking, and talking). Recently, normative percentile values (or "nomograms") of total gray matter volume as a function of age have been used as a potential reference application in clinical and research settings for elderly adults (Nobis et al., 2019).

4.2 Dynamic Association of Brain Development with Cognitive Development Growth Percentiles

While percentiles can help describe and put in perspective individual brain development, the cognitive impact of for example, being placed in the lowest 5th percentile remains unknown. As children develop and attain more skills, their brain structure (e.g., Marrus et al., 2018) and network functional connectivity at rest changes (e.g., Bruchhage et al., 2020). In order to link brain tissue volumes to cognitive development, we used a linear varying coefficient model with scores of overall cognitive functioning as the response and pGM, pWM and pCSF as predictors. Because of the broad age range of our sample, we had to use two different assessments for overall cognitive function. For ages two to six, we used the early learning coefficient, as well as the verbal and nonverbal developmental quotients from the Mullen Scale of Learning, while for children aged six and up, we used the full scale IQ of the WISC (Figure 8). We then identified intervals where the slopes of the varying coefficient models were found to be significantly away from zero, thus identifying time zones of significant positive and negative influence of cognitive scores on brain volume. In the younger group, higher ELC and nonverbal NVDQ were associated with higher pWM between three and a half and over four years, and lower pCSF from ages three and a half to four and a half. While time windows differ between the ELC and NVDQ with the latter being broader, it seems likely that the effect on ELC was driven by the effect of NVDQ as ELC is a summary

measure that combines VDQ and NVDQ. When investigating the influence of the FSIQ as a measure of overall function, we again showed increases in pWM to be associated with increases in FSIQ between ages nine to ten, while increases in FSIQ are associated with decreases in pCSF around age seven and ages nine to ten. Thus, increases in overall and nonverbal cognitive functioning are associated with the strength of dynamics that is the opposite for pWM and pCSF (Figure 8). Investigating the impact of brain development on cognitive overall functioning may then allow for more appropriate neurodevelopmental burden estimates and also help to identify primary risk factors in addition to objective and quantitative measures for assessing possible age-specific intervention impact at the individual and population level.

4.3 Individual Trajectory Modeling

Taking our approach one step further, we decided to reconstruct the individual smooth underlying trajectories at the subject level using the PACE approach, which we illustrated for three randomly selected boys and girls (Figures 9–10 respectively; Figures S.6–S.7 for raw volumes). Interestingly, even when longitudinal time points are extremely sparse, the fitted trajectories align well with the measurements. When using relatively highly sampled longitudinal data for one boy and girl (5 and 8 scans respectively), then using only the first 50% of their measurements for fitting the trajectories, future measurements were mostly contained within the confidence band of the fitted trajectories (Figure 11; Figure S.8 for raw volumes).

Placing individual participant data on growth curves can inform about both typical but also atypical development, including early detection of neurodevelopmental disorders or later developmental struggles, which in turn can make school entry more difficult, leading to more academic struggles in the future. For example, if a child's data would place it into the 5th percentile of pWM during a time window that has been identified to be sensitive to the establishment of overall cognitive functioning, follow up tests and if needed early interventions could help prevent possible future struggles from manifesting. The prediction of future neurocognive development using percentiles as a population-based reference has the potential to inform or predict whether a child might "fall off the curve".

4.4 Limitations

The PACE method proposed for the reconstruction of smooth trajectories from sparsely observed longitudinal data relies on the fact that the individual measurements are taken at random times. The estimation of individuals' trajectories, which relies on the estimated FPC scores, is implicitly conditional on the random observation times. The estimated FPC scores target the conditional expectation of the true scores, which given an individual's observations, is the best linear predictor (best predictor under Gaussianity) of the true scores, and this property does not require Gaussianity. The uniform confidence bands proposed for the prediction of reconstruction of the individual trajectories are however valid only under the assumption of Gaussianity of the true FPC scores.

We further note that the brain-for-age curves that we introduce in this paper may not represent the brain development in early childhood for the entire population of typically developing children in the US, or even in Providence, RI, due to the limited number of children involved in this study. Our main goal is rather to provide and illustrate a method to construct brain-for-age curves for neurodevelopmental studies.

4.5 Conclusions

We demonstrate that the PACE method is suitable to model and visualize trajectories of gray matter, white matter and cerebrospinal fluid development in sparse longitudinal data in a large paediatric cohort spanning early infancy to late childhood. These trajectories can be enriched by dynamic percentiles for sparsely measured brain tissue data (Figure 7) and applied at both population (Figures 4–8) and subject (Figures 9–11) levels. While we have been using this novel analysis method to model neuroanatomical development, it can be applied more broadly to any kind of sparse longitudinal data, including diffusion tensor imaging, functional and other magnetic resonance imaging data outputs.

5 Acknowledgement

The research was supported by NIH grant UG3-0D023313 (ECHO Program).

References

Aggarwal, O. P. (1955). Some minimax invariant procedures for estimating a cumulative distribution function. *Annals of Mathematical Statistics*, 26(3):450–463.

Avants, B. B., Tustison, N. J., Stauffer, M., Song, G., Wu, B., and Gee, J. C. (2014). The Insight ToolKit image registration framework. *Frontiers in Neuroinformatics*, 8:44.

Barnea-Goraly, N., Menon, V., Eckert, M., Tamm, L., Bammer, R., Karchemskiy, A., Dant, C. C., and Reiss, A. L. (2005). White matter development during childhood and adolescence: a cross-sectional diffusion tensor imaging study. *Cerebral Cortex*, 15(12):1848–1854.

Blakemore, S.-J. and Choudhury, S. (2006). Development of the adolescent brain: implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry*, 47(3–4):296–312.

Bondell, H. D., Reich, B. J., and Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika*, 97(4):825–838.

Brain Development Cooperative Group (2012). Total and regional brain volumes in a population-based normative sample from 4 to 18 years: the NIH MRI Study of Normal Brain Development. *Cerebral Cortex*, 22(1):1–12.

Bray, S., Krongold, M., Cooper, C., and Lebel, C. (2015). Synergistic effects of age on patterns of white and gray matter volume across childhood and adolescence. *eNeuro*, 2(4).

Bruchhage, M. M., Ngo, G.-C., Schneider, N., D'Sa, V., and Deoni, S. C. (2020). Functional connectivity correlates of infant and early childhood cognitive development. *Brain Structure and Function*, 225(2):669–681.

Cai, Z., Fan, J., and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95(451):888–902.

Carroll, C., Gajardo, A., Chen, Y., Dai, X., Fan, J., Hadjipantelis, P. Z., Han, K., Ji, H., Müller, H.-G., and Wang, J.-L. (2020). *fdapace: Functional Data Analysis and Empirical*

Dynamics. R package version 0.5.3, available at https://CRAN.R-project.org/package=fdapace.

Chen, Y., Gajardo, A., Fan, J., Zhong, Q., Dubey, P., Han, K., Bhattacharjee, S., and Hans-Georg, M. (2020). *frechet: Statistical Analysis for Random Objects and Non-Euclidean Data*. R package version 0.1.0, available at https://CRAN.R-project.org/package=frechet.

Cheng, C. and Parzen, E. (1997). Unified estimators of smooth quantile and quantile density functions. *Journal of Statistical Planning and Inference*, 59(2):291–307.

Cole, T. (1988). Fitting smoothed centile curves to reference data. *Journal of the Royal Statistical Society: Series A*, 151(3):385–406.

Cole, T. (1994). Growth charts for both cross-sectional and longitudinal data. *Statistics in Medicine*, 13:2477–2492.

Dai, X., Hadjipantelis, P., Wang, J.-L., Deoni, S. C., and Müller, H.-G. (2019). Longitudinal associations between white matter maturation and cognitive development across early childhood. *Human Brain Mapping*, 40:4130–4145.

Dean, D. C., Dirks, H., O'Muircheartaigh, J., Walker, L., Jerskey, B. A., Lehman, K., Han, M., Waskiewicz, N., and Deoni, S. C. (2014). Pediatric neuroimaging using magnetic resonance imaging during non-sedated sleep. *Pediatric Radiology*, 44(1):64–72.

Falk, M. (1983). Relative efficiency and deficiency of kernel type estimators of smooth distribution functions. *Statistica Neerlandica*, 37(2):73–83.

Falk, M. (1984). Relative deficiency of kernel type estimators of quantiles. *Annals of Statistics*, 12(1):261–268.

Gennatas, E. D., Avants, B. B., Wolf, D. H., Satterthwaite, T. D., Ruparel, K., Ciric, R., Hakonarson, H., Gur, R. E., and Gur, R. C. (2017). Age-related effects and sex differences in gray matter density, volume, mass, and cortical thickness from childhood to young adulthood. *Journal of Neuroscience*, 37(20):5065–5073.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., Evans, A. C., and Rapoport, J. L. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nature Neuroscience*, 2(10):861–863.

Giedd, J. N. and Rapoport, J. L. (2010). Structural MRI of pediatric brain development: what have we learned and where are we going? *Neuron*, 67(5):728–734.

Gilmore, J. H., Knickmeyer, R. C., and Gao, W. (2018). Imaging structural and functional brain development in early childhood. *Nature Reviews Neuroscience*, 19(3):123.

Giorgio, A., Watkins, K. E., Chadwick, M., James, S., Winmill, L., Douaud, G., De Stefano, N., Matthews, P. M., Smith, S. M., Johansen-Berg, H., et al. (2010). Longitudinal changes in grey and white matter during adolescence. *NeuroImage*, 49(1):94–103.

Gogtay, N. and Thompson, P. M. (2010). Mapping gray matter development: implications for typical development and vulnerability to psychopathology. *Brain and Cognition*, 72(1):6–15.

Hall, P., Müller, H.-G., and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Annals of Statistics*, 34(3):1493–1517.

He, X. (1997). Quantile curves without crossing. *The American Statistician*, 51(2):186–192.

Hendricks, W. and Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association*, 87(417):58–68.

Huang, J. Z., Wu, C. O., and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica*, 14(3):763–788.

Kloeckner, B. (2010). A geometric study of Wasserstein spaces: Euclidean spaces. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 9(2):297–323.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46:33–50.

Koenker, R., Ng, P., and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, 81:673–680.

Lancaster, J. L., Tordesillas-Gutiérrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., Mazziotta, J. C., and Fox, P. T. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping*, 28(11):1194–1205.

Lebel, C. and Beaulieu, C. (2011). Longitudinal development of human brain wiring continues from childhood into adulthood. *Journal of Neuroscience*, 31(30):10937–10947.

Lebel, C., Walker, L., Leemans, A., Phillips, L., and Beaulieu, C. (2008). Microstructural maturation of the human brain from childhood to adulthood. *NeuroImage*, 40(3):1044–1055.

Leblanc, A. (2012). On estimating distribution functions using Bernstein polynomials. *Annals of the Institute of Statistical Mathematics*, 64(5):919–943.

Marrus, N., Eggebrecht, A. T., Todorov, A., Elison, J. T., Wolff, J. J., Cole, L., Gao, W., Pandey, J., Shen, M. D., Swanson, M. R., et al. (2018). Walking, gross motor development, and brain functional connectivity in infants and toddlers. *Cerebral Cortex*, 28(2):750–763.

Matsuzawa, J., Matsui, M., Konishi, T., Noguchi, K., Gur, R. C., Bilker, W., and Miyawaki, T. (2001). Age-related volumetric changes of brain gray and white matter in healthy infants and children. *Cerebral Cortex*, 11(4):335–342.

Mullen, E. M. (1995). *Mullen Scales of Early Learning*. Circle Pines, MN: American Guidance Service.

Müller, H.-G. and Stadtmüller, U. (1999). Multivariate boundary kernels and a continuous least squares principle. *Journal of the Royal Statistical Society: Series B*, 61:439–458.

Nobis, L., Manohar, S. G., Smith, S. M., Alfaro-Almagro, F., Jenkinson, M., Mackay, C. E., and Husain, M. (2019). Hippocampal volume across age: Nomograms derived from over 19,700 people in UK Biobank. *NeuroImage: Clinical*, 23:101904.

O'Muircheartaigh, J., Dean III, D. C., Ginestet, C. E., Walker, L., Waskiewicz, N., Lehman, K., Dirks, H., Piryatinsky, I., and Deoni, S. C. (2014). White matter development and early cognition in babies and toddlers. *Human Brain Mapping*, 35(9):4475–4487.

Panaretos, V. M. and Zemel, Y. (2016). Amplitude and phase variation of point processes. *Annals of Statistics*, 44(2):771–812.

Parzen, E. (1979). Nonparametric statistical data modeling. *Journal of the American Statistical Association*, 74(365):105–121.

Petersen, A. and Müller, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *Annals of Statistics*, 44(1):183–218.

Petersen, A. and Müller, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Annals of Statistics*, 47(2):691–719.

Read, R. (1972). The asymptotic inadmissibility of the sample distribution function. *Annals of Mathematical Statistics*, 43(1):89–95.

Rigby, R. A. and Stasinopoulos, D. M. (2004). Smooth centile curves for skew and kurtotic data modelled using the Box–Cox power exponential distribution. *Statistics in Medicine*, 23(19):3053–3076.

Samanta, M. (1989). Non-parametric estimation of conditional quantiles. *Statistics & Probability Letters*, 7(5):407–412.

Şentürk, D. and Müller, H.-G. (2008). Generalized varying coefficient models for longitudinal data. *Biometrika*, 95(3):653–666.

Şentürk, D. and Müller, H.-G. (2010). Functional varying coefficient models for longitudinal data. *Journal of the American Statistical Association*, 105(491):1256–1264.

Şentürk, D. and Nguyen, D. V. (2011). Varying coefficient models for sparse noise-contaminated longitudinal data. *Statistica Sinica*, 21(4):1831.

Silbereis, J. C., Pochareddy, S., Zhu, Y., Li, M., and Sestan, N. (2016). The cellular and molecular landscapes of the developing human central nervous system. *Neuron*, 89(2):248–268.

Sturm, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces (Paris, 2002)*, 338:357–390.

Toga, A. W., Thompson, P. M., and Sowell, E. R. (2006). Mapping brain maturation. *Focus*, 29(3):148–390.

UNICEF, WHO, and World Bank (2012). Levels and trends in child malnutrition: UNICEF/WHO/The World Bank joint child malnutrition estimates. *UNICEF: New York, NY, USA*.

Vuoksimaa, E., Panizzon, M. S., Franz, C. E., Fennema-Notestine, C., Hagler, D. J., Lyons, M. J., Dale, A. M., and Kremen, W. S. (2018). Brain structure mediates the association between height and cognitive ability. *Brain Structure and Function*, 223(7):3487–3494.

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295.

Wechsler, D. (2012). Wechsler preschool and primary scale of intelligence—fourth edition. *The Psychological Corporation San Antonio, TX*.

WHO (2006). WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development.

Wu, Y., Fan, J., Müller, H.-G., et al. (2010). Varying-coefficient functional linear regression. *Bernoulli*, 16(3):730–758.

Yang, S.-S. (1985). A smooth nonparametric estimator of a quantile function. *Journal of the American Statistical Association*, 80(392):1004–1011.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.