NeuroImage 237 (2021) 118079

Contents lists available at ScienceDirect

NeuroImage

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

Modeling sparse longitudinal data in early neurodevelopment☆,☆☆,★

Yaqing Chen[a,1], Paromita Dubey[b,1], Hans-Georg Müller[a,1], Muriel Bruchhage[c,d], Jane-Ling Wang[a], Sean Deoni[c,d,e,f,*]

[a] Department of Statistics, University of California, Davis, Davis, CA, 95616, USA
[b] Department of Statistics, Stanford University, Stanford, CA, 94305, USA
[c] Advanced Baby Imaging Lab, Hasbro Children's Hospital, Rhode Island Hospital, Providence, RI, 02903, USA
[d] Department of Pediatrics, Warren Alpert Medical School at Brown University, Providence, RI, 02912, USA
[e] Department of Radiology, Warren Alpert Medical School at Brown University, Providence, RI, 02912, USA
[f] Maternal, Newborn, and Child Health Discovery & Tools, Bill & Melinda Gates Foundation, Seattle, WA, USA

ARTICLE INFO

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

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.

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 microstructure tissue maturation, including increasing myelination and white matter volume, changing cortical morphometry, and increasing subcortical gray matter volumes, synapse and neuronal density.

By scanning and assessing children throughout early development, longitudinal studies are in theory able to characterize brain growth patterns consistently with age, and investigate associations with current and future cognitive performance. In fact, efforts to translate developmental trajectories into growth curves have been initiated for neuropsychiatric disorders, believed to show abnormal neurodevelopmental origins and trajectories (Di Martino et al., 2014; Dong et al., 2020; Gur et al., 2014; Kessler et al., 2016). 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. By far the most commonly used approaches are built on mixed effects modeling, which features fixed effects to account for average population trends, and random effects to characterize individual

☆ 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.
★ RESONANCE Consortium
* Corresponding author at: Department of Pediatrics, Warren Alpert Medical School at Brown University, Providence, RI, 02912, USA.
E-mail address: sean_c_deoni@brown.edu (S. Deoni).
[1] Co-first authors.

https://doi.org/10.1016/j.neuroimage.2021.118079.
Received 26 July 2020; Received in revised form 9 April 2021; Accepted 12 April 2021
Available online 14 May 2021.

deviations from the mean trajectory (Bates et al., 2014; Bernal-Rusiel et al., 2013a; 2013b; Lindstrom and Bates, 1990; Pinheiro and Bates, 2020; Sanford et al., 2018).

Growth patterns are typically studied in the 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 World Health Organization (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 frequently 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. Foregoing 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

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

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 $\lambda \geq 0$. Koenker and Bassett (1978) impose a linearity condition $g(x) = x^\top \beta$. However, this method suffers from quantile crossing, i.e. quantile lines $\hat{g}_\tau$ 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 Bondell et al. (2010); He (1997).

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 (Bray et al., 2015; Gennatas et al., 2017; Giorgio et al., 2010; Gogtay and Thompson, 2010; Lebel and Beaulieu, 2011; Lebel et al., 2008; Tamnes et al., 2017; Yu et al., 2020). 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 characterize 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 often suboptimal. While nonlinear parametric approaches that rely on the Gompertz function, power law and Weibull models have been used to obtain normative growth curves (Peterson et al., 2018; Sadeghi et al., 2013), such models run the risk of being misspecified. There is therefore a need for more flexible nonparametric methods. Nonparametric methods for conditional quantile estimation have been proposed based on kernel smoothing and splines (e.g., Hendricks and Koenker, 1992; Samanta, 1989), 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) makes it possible to construct age dynamic growth percentiles at the population level. We use this novel approach to develop normative percentile "brain growth charts" for proportions of GM, WM and CSF volumes from infancy to pre-adolescence. We then examine dynamic associations between brain growth trajectories and cognitive scores derived from the Mullen Scales of Early Learning (Mullen, 1995) and Wechsler Intelligence Scale for Children (WISC) Fourth Edition (Wechsler, 2012) in early and late childhood. To eliminate confusion regarding the usage of the term *functional*, which also appears in a different context in neuroimaging, we note that the word "functional" in this paper refers to the statistical methods used for the analyses, which are from the branch of statistics called functional data analysis (FDA) that deals with methodology for data samples that are generated by underlying curves, and more general objects varying over a continuum.

Table 1
Distribution of numbers of repeated scans 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

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 with ages ranging from 0 to 10 years.

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

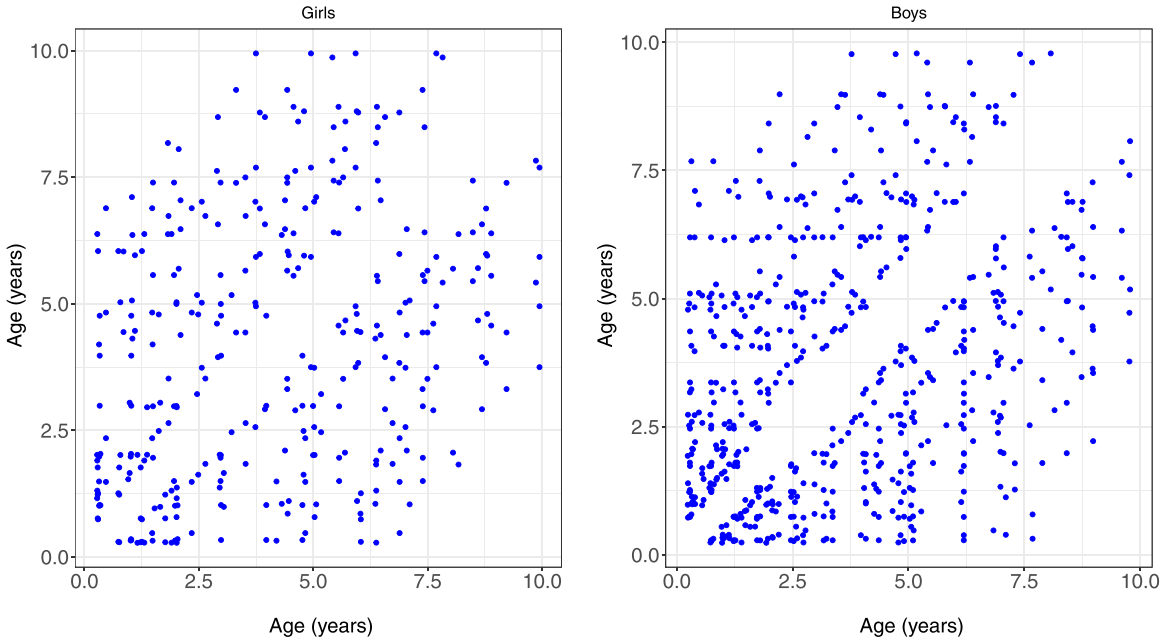

Fig. 1. Design plot displaying all pairs of ages for all included girls (left) and boys (right) from the RESONANCE cohort.

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 was 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 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 were 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 Fig. 1 illustrates all pairwise measurements with age differentiated by gender. For the RESONANCE cohort, this plot reveals the sparsity of the times when measurements were taken. Most of the children had only one scan (Table 1), and most of the measurements were taken at ages less than 5 years (Fig. 2).

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 $1.2 \times 1.2 \times 1.2$ mm^3, resampled to $0.9 \times 0.9 \times 0.9$ mm^3. Sequence specific parameters were: TE = 6.9 ms; TR = 16ms; inversion preparation time = 950 ms; flip angle = 15 degrees; BW = 450 Hz/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 age-specific anatomical T1-weighted templates were created corresponding to 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 42, 48, 60, 72, 84, 96 and 108 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 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 (Fig. 3, and Figs. S.1–S.8 in the Supplement).

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 functioning 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.

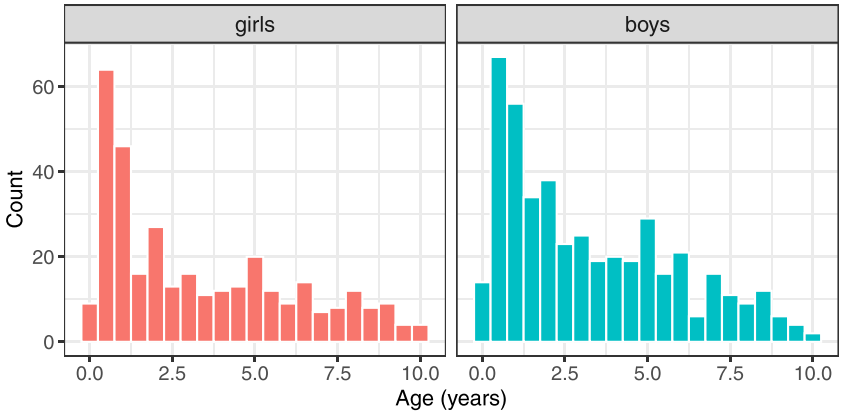

Fig. 2. Distributions of ages of observations for girls (left) and boys (right).

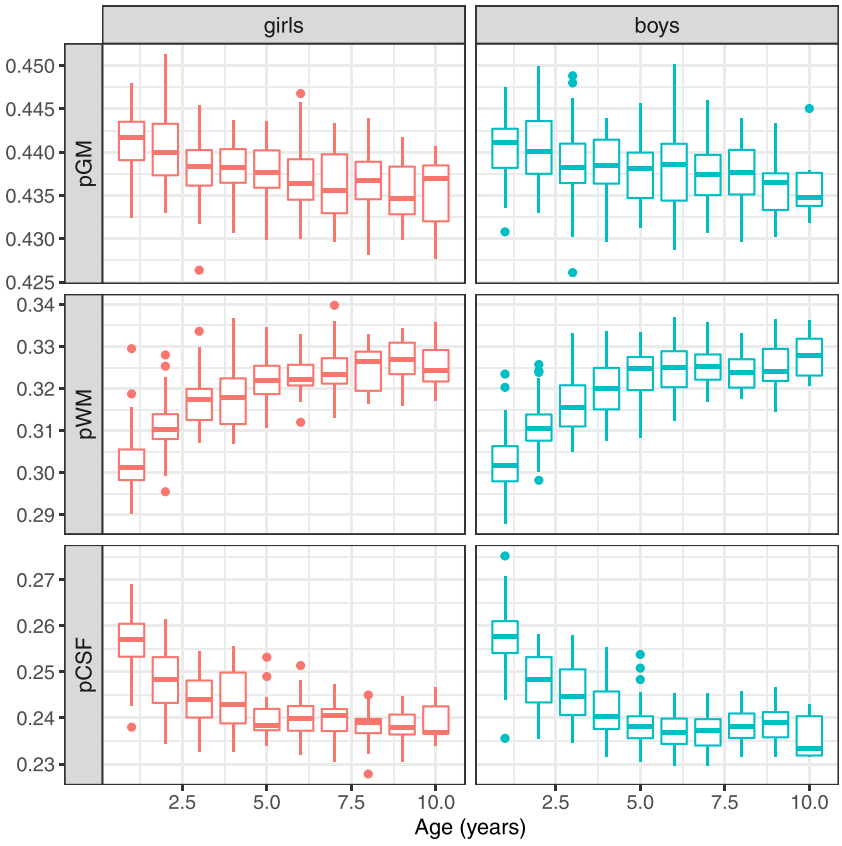

Fig. 3. Distributions of pGM (top), pWM (middle), and pCSF (bottom) observed within each year increment of ages for girls (left) and boys (right).

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 (Hall et al., 2006; Wang et al., 2016) has been devised for fully observed or densely observed curves but challenges arise when applying this approach in longitudinal settings when one has only very few repeated measurements for each subject or the measurements are irregularly timed. While the mean function can be estimated by smoothing across neighborhoods even in sparse settings as depicted in Fig. 1, the estimation of the covariance surface, which is the backbone of FPCA, is more complex in sparse settings (Wang et al., 2016; Yao et al., 2005). 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 with very sparse and irregular measurements are paramount. Previously, FPCA has been proposed for longitudinal neuroimaging data (Hyun et al., 2016) where the authors use spatio-temporal Gaussian process modeling, and in case of a sparse design, a linear random effects model to characterize deviations around the mean trajectory.

Principal component analysis through conditional expectation (PACE) (Yao et al., 2005) is specifically 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, under minimal assumptions one can get es-

timates 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 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 a functional concurrent regression (varying coefficient) model, as described in Section 2.5.2, to study the association of pGM, pWM and pCSF with cognitive scores. Various estimation techniques for these models are available (Cai et al., 2000; Dai et al., 2019; Huang et al., 2004; Şentürk and Müller, 2010; Wu et al., 2010).

To derive and construct conventional quantile growth charts for pGM, pWM and pCSF, as described in Section 2.5.3, we employ local Fréchet regression (Petersen and Müller, 2019), which is a non-parametric regression method for responses lying in metric spaces that are coupled with Euclidean predictors. We apply this approach for the case where the responses are distributions to 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

$$\mu(t) = \mathbb{E}(X(t)) \text{ and}$$

$$C(s,t) = \mathrm{Cov}(X(s), X(t)) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s)\phi_k(t),$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ 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

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

where $\xi_k = \int_{\mathcal{I}} (X(t) - \mu(t))\phi_k(t)dt$ are the functional principal components. The ξ_k are zero mean uncorrelated random variables, accounting for random fluctuations of the trajectory $X(t)$ around the mean curve $\mu(t)$. For the purpose of dimension reduction, the first K eigenfunctions are used so that $X(t) \approx \mu(t) + \sum_{k=1}^{K} \xi_k \phi_k(t)$ is represented using FPCs $(\xi_1, \ldots, \xi_K)$. We describe the details of the estimation steps for $\mu(t)$, $C(s,t)$, λ_k and ϕ_k in Section 5.1 of the Supplement. To determine the necessary smoothing bandwidths in an automatic data-adaptive way, for the RESONANCE data, we use the geometric mean of the bandwidth obtained using generalized cross validation and a data adaptive minimum cutoff (5% of the age domain for mean estimation and 10% of the age domain for covariance estimation) as the bandwidth choices for the mean and covariance smoothing steps. This automatic bandwidth choice is included as an option in the R package `fdapace`.

For visualization in FDA, modes of variation plots provide an insightful representation of variance decomposition for a given 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 kth mode of variation is defined as the family of functions

$$M_{k,s}(t) = \mu(t) \pm s\sqrt{\lambda_k}\phi_k(t), \ t \in \mathcal{I}, \ s \in \mathbb{R}, \tag{2}$$

where the above population quantities are estimated using their sample counterparts, as described in Section 5.1 of the Supplement, and the modes of variation are typically visualized as functions in t over a grid of points $s \in [-2, 2]$.

2.5.2. Varying coefficient modeling

The functional concurrent regression model between functional responses $Y(t)$ and functional predictors $X(t)$ is given by

$$\mathbb{E}(Y(t)|X(t)) = \beta_0(t) + \beta(t)X(t), \ t \in \mathcal{I}, \tag{3}$$

where $\beta_0(t)$ and $\beta(t)$ are smooth coefficient functions. 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; 2004; Şentürk and Müller, 2008; 2010; Şentürk and Nguyen, 2011; Wu et al., 2010). To assess the strength of the association, one can use the time-varying R^2 function

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

Larger values of $R^2(t)$ indicate that the model explains more of the variability in the response $Y(t)$ at time t. The sign of the slope function $\beta(t)$ at time t indicates whether the association between $Y(t)$ and $X(t)$ tends to be positive or negative.

2.5.3. Dynamic quantile modeling using local Fréchet regression

To obtain age-specific quantile curves for brain modalities such as pGM, pWM, and pCSF, we essentially need to estimate their age-dependent distributions. We address this by introducing an approach that utilizes local Fréchet regression with distributions as response and age as predictor, where we use the $\mathcal{L}^2$ Wasserstein metric in the distribution space.

Given a closed interval $D \subset \mathbb{R}$, which is the domain of the variable of interest, we consider 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

$$d_W(q_1, q_2) = \left\{ \int_0^1 \left[F_1^{-1}(u) - F_2^{-1}(u)\right]^2 \mathrm{d}u \right\}^{1/2},$$
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 the distribution 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, $F^{-1}(u) = \inf\{x \in D : F(x) \geq u\}$, for $u \in (0, 1)$.

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

$$\nu_{\oplus}(t) = \mathrm{argmin}_{q \in \mathcal{W}} M(q, t).$$

Given the distribution $\nu_{\oplus}(t)$, the τth conditional quantile of P given $T = t$ can be expressed as $g_\tau(t) = F^{-1}_{\nu_{\oplus}(t)}(\tau)$, for $\tau \in (0, 1)$, where $F^{-1}_{\nu_{\oplus}(t)}$ is the quantile function of $\nu_{\oplus}(t)$. Thus, in order to estimate the conditional quantiles, which then lead to the percentile curves across t, it suffices to estimate the conditional Fréchet means $\nu_{\oplus}(t)$ for $t \in \mathcal{I}$. To this end, we employ the local Fréchet regression approach (Petersen and Müller, 2019), as follows.

Suppose $\{(T_i, P_i)\}_{i=1}^n$ are independent realizations of (T, P). 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 suitable representations of the distributions P_i, which include cumulative distribution functions (cdf) (e.g., Aggarwal, 1955; Falk, 1983; Leblanc, 2012; Read, 1972), quantile functions (e.g., Cheng and Parzen, 1997; Falk, 1984; Parzen, 1979;

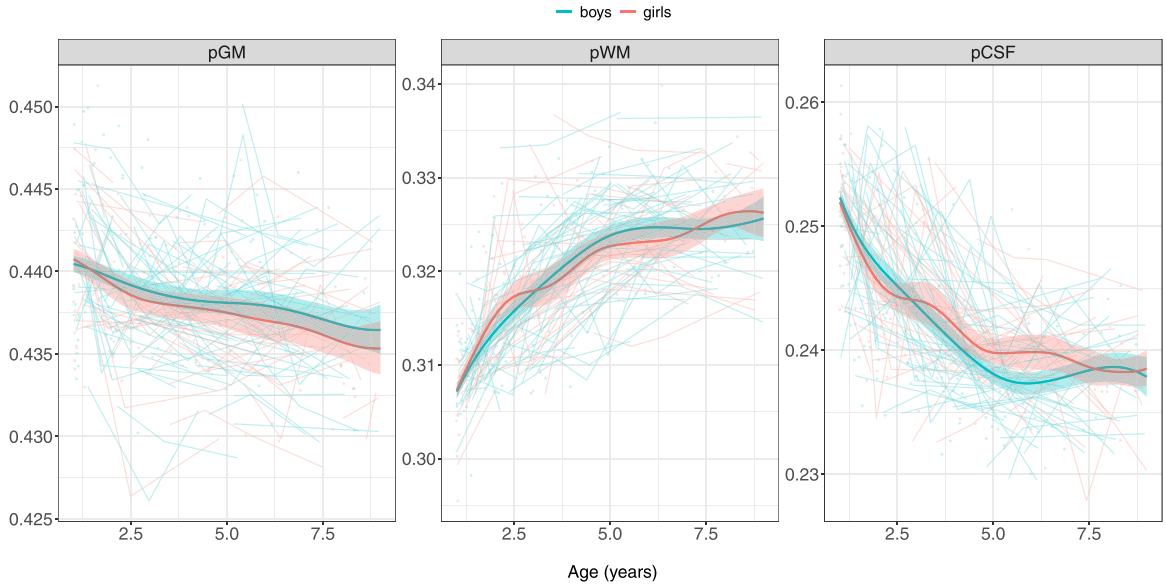

Fig. 4. Population level mean functions (dark solid lines) for proportion of gray matter (left), white matter (middle) and CSF (right) evolution for children in the RESONANCE cohort. The light ribbons correspond to 95% pointwise confidence intervals and the background points and lines to the underlying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

Yang, 1985) or density functions (e.g., Panaretos and Zemel, 2016; Petersen and Müller, 2016) of the underlying distributions. For any distribution $q \in \mathcal{W}$, we denote the estimated distribution by $\tilde{q} = q(\tilde{F})$, where $\tilde{F}$ is an estimate of the cdf that is 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 $\tilde{P}_i$, a data-based local Fréchet regression estimate is

$$\tilde{\ell}_{\oplus}(t) = \mathrm{argmin}_{q\in\mathcal{W}} n^{-1} \sum_{i=1}^{n} \hat{w}(T_i, t, h) d_W^2(\tilde{P}_i, q).$$

Here, $\hat{w}(s,t,h) = \hat{\sigma}_0^{-2} H_h(s-t)[\hat{\kappa}_2 - \hat{\kappa}_1(s-t)]$, where $H_h(\cdot) = H(\cdot/h)/h$, $H(\cdot)$ is a smoothing kernel, i.e., a density function symmetric around zero, and $h = h(n) > 0$ is a bandwidth sequence, $\hat{\kappa}_z = n^{-1} \sum_{i=1}^{n} H_h(T_i - t)(T_i - t)^z$, for $z = 0, 1, 2$, and $\hat{\sigma}_0^2 = \hat{\kappa}_0\hat{\kappa}_2 - \hat{\kappa}_1^2$.

However, in the RESONANCE data, only one measurement is available per distribution at most ages and therefore the above distribution estimation methods cannot be applied directly. To address this, we divide the age domain $\mathcal{I}$ into bins $S_1 = [a_0, a_1), S_2 = [a_1, a_2), \ldots, S_B = [a_{B-1}, a_B]$, with $\min(\mathcal{I}) = a_0 < a_1 < \ldots < a_B = \max(\mathcal{I})$. Pooling the observations from all the subjects, denoting the paired observations by (t_j, Y_j), where t_j is the time of the measurement and Y_j is the associated measurement generated from a random distribution $P(t_j)$ at age t_j, for $j = 1, \ldots, m$, we then obtain for the midpoint $b_k = (a_{k-1} + a_k)/2$ of the kth bin an estimate $\hat{P}(b_k)$ of the distribution $P(b_k)$ that is based on the observations $\{Y_j : t_j \in S_k\}$ using those data associated with ages t_j that fall into the bin S_k. Then the local Fréchet regression estimate of the distribution $\nu_{\oplus}(t)$ for the RESONANCE data becomes

$$\hat{\ell}_{\oplus}(t) = \mathrm{argmin}_{q\in\mathcal{W}} \sum_{k=1}^{B} \hat{w}(b_k, t, h) d_W^2(\hat{P}(b_k), q). \quad (4)$$

In a next step, for any $\tau \in (0,1)$, an estimate for the τth conditional quantile curve $g_\tau(\cdot)$ is obtained as

$$\hat{g}_\tau(t) = F^{-1}_{\hat{\ell}_{\oplus}(t)}(\tau) \quad \text{for all } t \in \mathcal{I}, \quad (5)$$

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

The choice of the number of bins, B affects the balance between bias and variance. To obtain reasonable preliminary estimates $\hat{P}(b_k)$ of the distributions $P(b_k)$, we need a sufficient number of measurements within each bin. However, when B is too small, observations which are not very close to the midpoints b_k of each bin are involved in constructing the preliminary estimates $\hat{P}(b_k)$, potentially incurring some bias for the estimates $\hat{P}(b_k)$. On the other hand, the set of the midpoints of the bins needs to be a relatively dense grid on the age domain so that local Fréchet regression method can be applied. Balancing these considerations, for the RESONANCE data, we set the number of bins at $B = 35$, where the bins are constructed such that each bin contains about 20 measurements. The bandwidth h in (4) was chosen by 10-fold cross validation simultaneously for pGM, pWM, and pCSF, with details described in Section S.3 of the Supplement. Specifically, the bandwidth used for the analysis of proportions of GM, WM, and CSF is 1.65.

3. Results

3.1. Population level analysis

3.1.1. Mean function

Figure 4 illustrates the population level mean curves for the development of the proportions of pGM, pWM and pCSF relative to total brain volume 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 bootstrap replicates of the mean function and then obtained the pointwise cutoffs at the 95% level. No major differences were detected in the levels of the mean proportions between boys and girls (Fig. 4). However, there are significant differences in the mean raw brain volume levels between boys and girls as illustrated in Fig. S.9 in the Supplement.

Figure 5 shows the first two dominant eigenfunctions and Fig. 6 the corresponding first two modes of variation obtained using the PACE approach 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 with progressing age. 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 func-

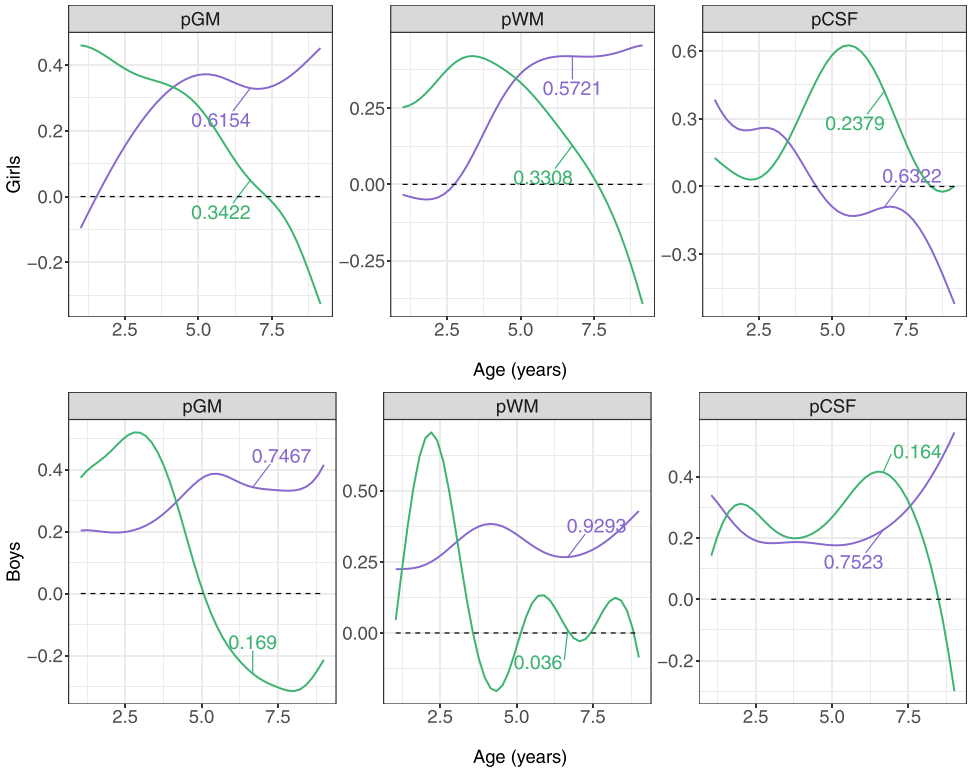

Fig. 5. The first two eigenfunctions for FPCA of proportion of gray matter (left), white matter (middle) and CSF (right) for girls (top panel) and boys (bottom panel) in the RESONANCE cohort. The numbers corresponding to the curves indicate the fraction of variance explained by each of the eigenfunctions.

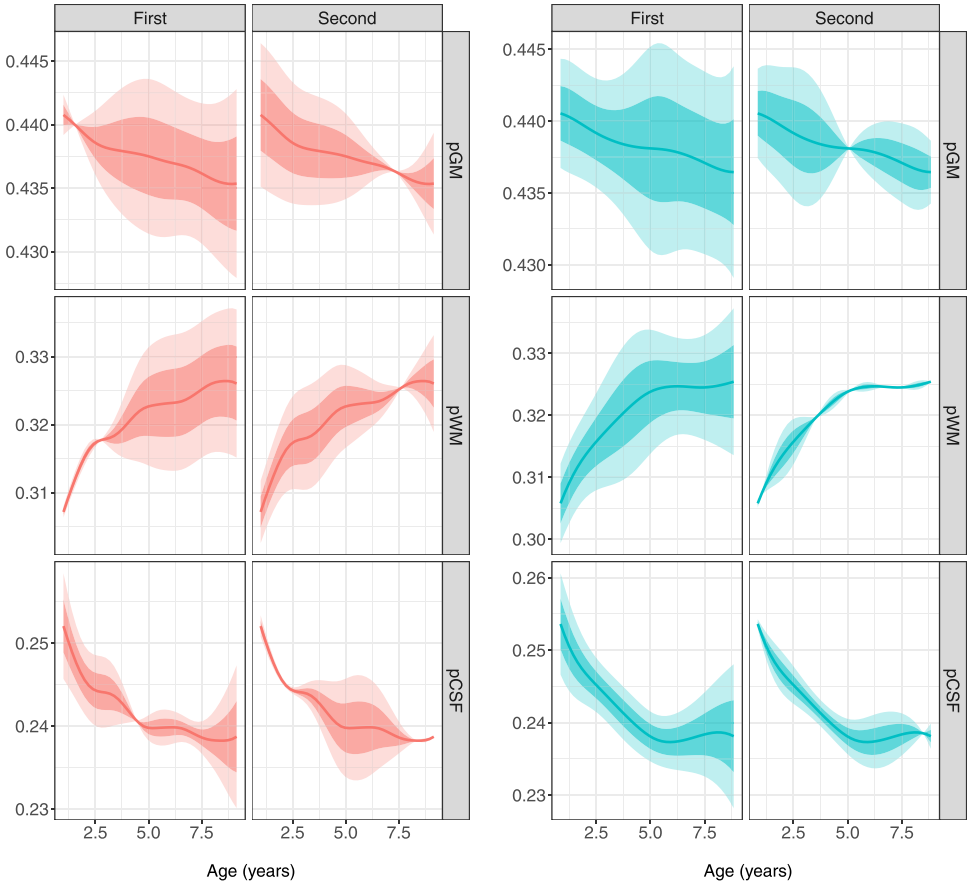

Fig. 6. First and second modes of variation plot for proportion of gray matter (top), proportion of white matter (middle) and proportion of CSF (bottom) for girls (left) and boys (right) in the RESONANCE cohort. The solid line corresponds to the mean function, the boundaries of the darker ribbon correspond to the kth mode of variation $M_{k,s}(t)$ with $s \in \{-1, 1\}$ as per (2) and the boundary of the lighter ribbon corresponds to $M_{k,s}(t)$ with $s \in \{-2, 2\}$ as per (2), where $k = 1$ and 2 correspond to the first and second modes of variation, respectively.

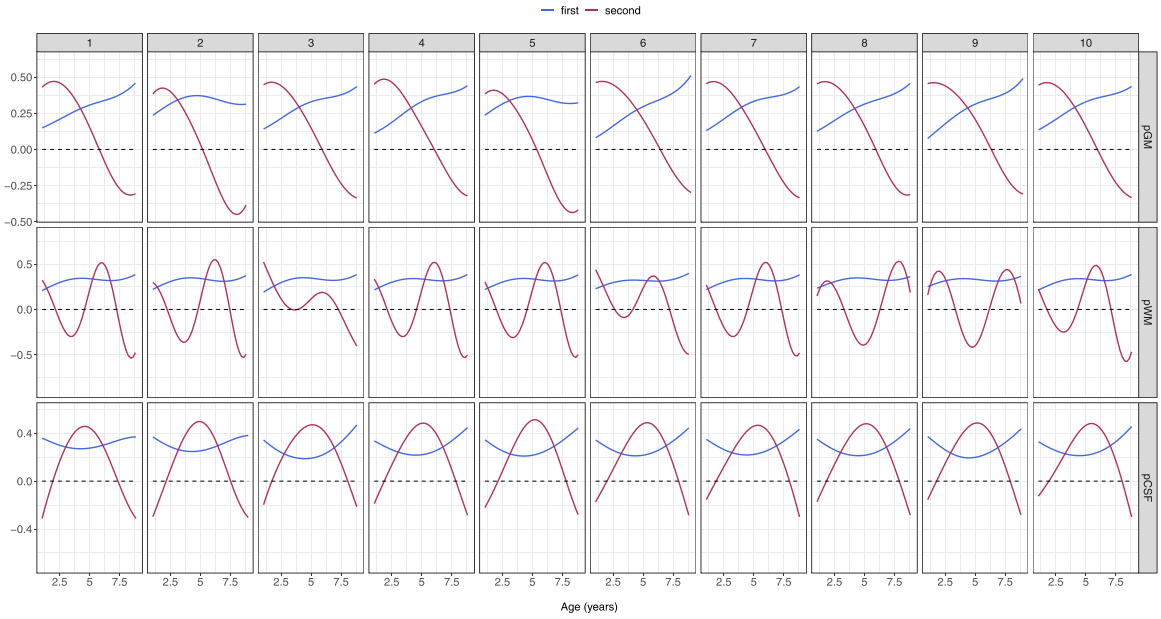

Fig. 7. Eigenfunction stability across reduced samples for boys.

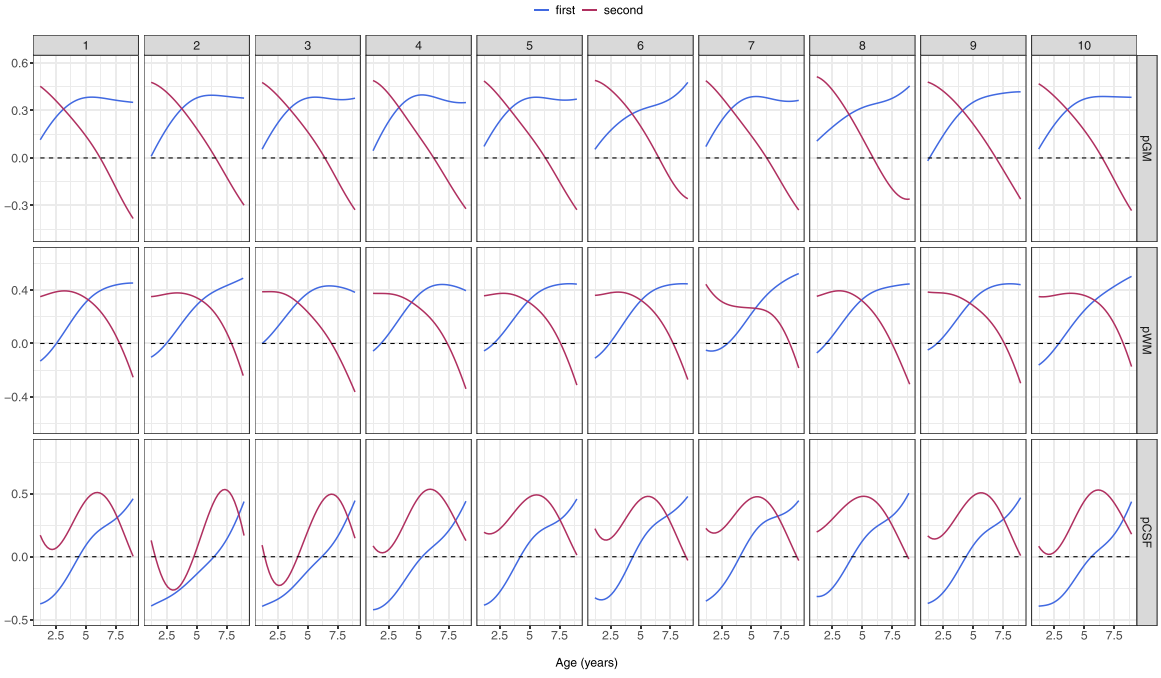

Fig. 8. Eigenfunction stability across reduced samples for girls.

tion as reflected in the corresponding first mode of variation in Fig. 6. The contrasting shapes between different growth periods form the second mode of variation for pGM and pWM in girls and also for pCSF in boys. Figures S.10 and S.11 in the Supplement show the corresponding behavior for the raw volumes.

For reliable interpretation of the shapes of the eigenfunctions, one must evaluate the in sample stability of the eigenfunction estimates. For this purpose we take subsamples, leaving out 10% of the subjects randomly every time. For each subsample, we estimate the covariance surface and the corresponding eigenfunctions. In Figs. 7 and 8 we illustrate that the estimated eigenfunctions are quite stable across 10 random subsamples, both for boys and girls. We also observe that the larger the fraction of variance explained by an eigenfunction, the greater is its stability. For example, the first eigenfunction of pWM for boys explains about 93% and the second explains about 3.6% in the actual data. In Fig. 7 we see that while the first eigenfunction for pWM is very stable, the second is less stable across the subsamples.

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 Fig. 9 (see Fig. S.12 in the Supplement for a version with raw data overlaid). 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 distribu-

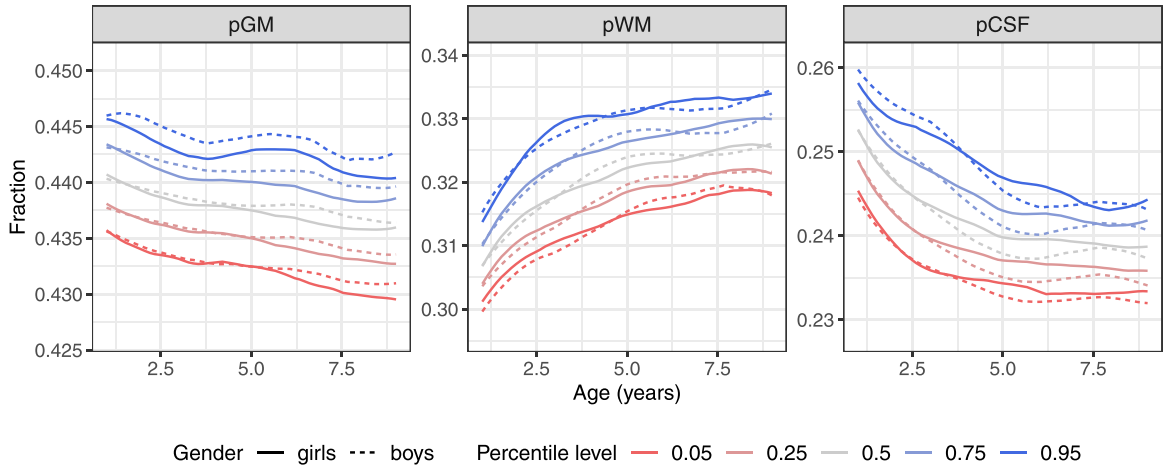

Fig. 9. Age-based dynamic percentiles of proportions of GM (pGM, left), WM (pWM, middle) and CSF (pCSF, right) estimated by local Fréchet regression as per (5) for the RESONANCE data.

tions in Fig. 3. Fluctuation in the dynamic percentile curves of pCSF along with age 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 Fig. S.13 and total brain volumes (TBV) in Fig. S.14 in the Supplement. 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 2.5.2 under statistical methods, with the time-varying ELC, VDQ, NVDQ and FSIQ scores as the response and pGM, pWM and pCSF as predictors. Since ELC, VDQ and NVDQ scores are early learning indicators (Mullen, 1995), they are available only for infants, and therefore we study these scores in the age interval 2 to 5.5 years. FSIQ scores, on the other hand, are indicating cognitive development in later ages, so we study the association of FSIQ scores with brain development in the age interval 6 to 10 years. Since the cognitive scores in early and late childhood are very different from each other, we did not put the association across all ages in one plot in Fig. 10. Another reason is that the observations used to obtain the plots for the age interval 2 to 5.5 years are disjoint from the observations used in the plot for the age interval 6 to 10 years. Due to the extreme sparsity, we do not differentiate this analysis across genders.

Figure 10 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. We note that these inferences are based on pointwise not uniform confidence bands and thus are not adjusted for multiple comparisons, and therefore should be viewed as exploratory and more indicative than conclusive.

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 illustrate this approach for predicting the longitudinal evolution of pGM, pWM and pCSF for three randomly selected boys in Fig. 11 and three randomly selected girls in Fig. 12. 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 Figs. S.15 and S.16 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. Fig. 13 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, even though they clearly carry additional random aberrations from their target values; these are reflected in our model as random errors. 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 Fig. S.17 in the Supplement. For the randomly selected boys, future measurements for the raw volumes are found to lie outside of the confidence bands, yet overall quite close to the fitted trajectory. For a quantitative check, we consider only the first two scans for all subjects having 3 or more scans and obtain their fitted trajectories. Table 2 illustrates that the fits obtained using PACE have a lower mean squared error compared to using just the group mean.

To further illustrate the utility of the PACE approach, we compared our findings with those of the LME models and the NLME models, which we implemented using the R package `nlme` (Pinheiro et al., 2020). Specifically, for a sample of observations from subjects $i = 1, \dots, n$, $\{(t_{ij}, X_{ij}) : j = 1, \dots, n_i\}_{i=1}^{n}$, we consider two LME models as follows:

$$X_{ij} = \beta_0 + \beta_1 t_{ij} + \alpha_{0i} + \epsilon_{ij}, \tag{6}$$

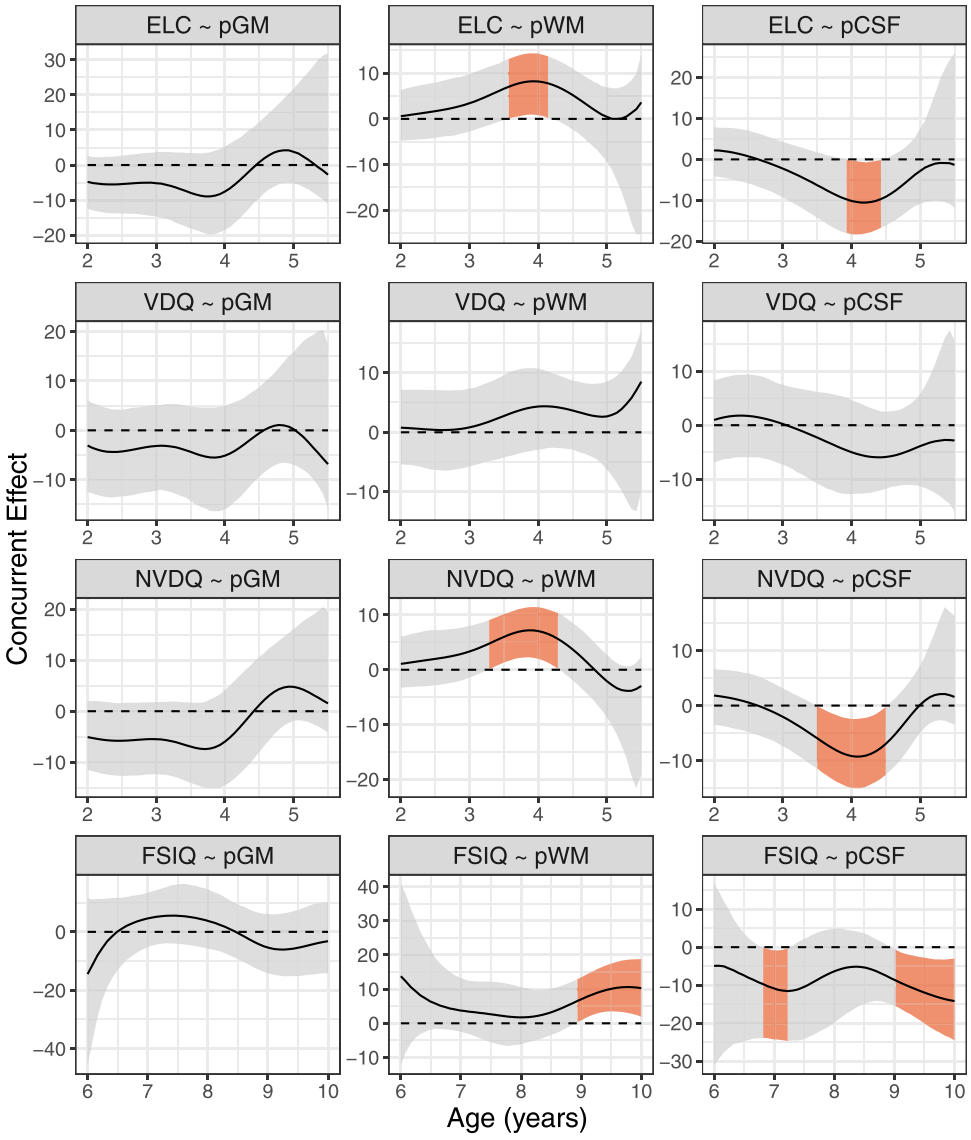

Fig. 10. The regression slope function (black solid line) along with 95% pointwise confidence intervals (the grey ribbon) for varying coefficient model in equation (3) of Section 2.5.2 with ELC, VDQ, NVDQ and FSIQ as response (top to bottom) and pGM (left), pWM (middle) and pCSF (right) as predictors. The orange portions of the ribbon indicate intervals where the slope function is significantly away from zero, based on pointwise unadjusted analysis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
MSE/10^{-5} for fitted individuals' trajectories by using PACE with only first two scans for individuals comprising of more than 2 scans (with the minimum numbers of FPCs with at least 95% FVE) as compared to the group mean trajectory.

	Girls			Boys		
	pGM	pWM	pCSF	pGM	pWM	pCSF
PACE	0.74	3.45	2.09	0.98	1.75	2.22
Mean	1.41	3.51	2.29	2.01	3.75	2.59

$$X_{ij} = \beta_0 + \alpha_{0i} + (\beta_1 + \alpha_{1i})t_{ij} + \epsilon_{ij}, \tag{7}$$

where β_0 and β_1 are fixed effects, and α_{0i} and α_{1i} are random intercepts and random slopes, respectively. For NLME modeling, inspired by the shape of the population level mean functions estimated by the PACE approach, we fitted the following NLME models:

$$X_{ij} = \beta_0 + \beta_1 \exp(\beta_2 t_{ij}) + \alpha_{0i} + \epsilon_{ij}, \tag{8}$$

$$X_{ij} = \beta_0 + \alpha_{0i} + (\beta_1 + \alpha_{1i}) \exp((\beta_2 + \alpha_{2i})t_{ij}) + \epsilon_{ij}, \tag{9}$$

where $\{\beta_l\}_{l=0}^2$ are fixed effects, and $\{\alpha_{li}\}_{l=0}^2$ are random effects, respectively. We found that the algorithm failed to fit model (9) for the RESONANCE data. The estimates of the population level mean functions (i.e., the fixed effects for LME and NLME models) for girls and boys obtained by the LME and NLME models and the PACE approach are shown in Figs. S.18 and S.19 in the Supplement, respectively, where the raw observations are overlaid.

To compare the performances of these different approaches, we computed the average subject-specific mean squared errors (ASMSEs),

$$\text{ASMSE} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}(\hat{X}_{ij} - X_{ij})^2. \tag{10}$$

The results are summarized in Table 3, where for the PACE approach, the fits are obtained based on the minimum numbers of FPCs with at least 95% FVE, i.e., $\hat{X}_{ij} = \hat{\mu}(t_{ij}) + \sum_{k=1}^{K}\hat{\xi}_{ik}\hat{\phi}_k(t_{ij})$ with $K = 2$ for pGM for girls and pWM for boys and $K = 3$ for the others, and $\hat{\xi}_{ik}$ as per (11). The results clearly demonstrate that the PACE approach yields considerably better fits to the data in all six cases considered ({pGM, pWM, pCSF} × {girls, boys}) than all the LME and NLME models. This is owing to the flexibility of the PACE approach as a nonparametric method, where no constraint is imposed on the shape of the mean and the covariance structure, as opposed to parametric methods like LME and NLME models.

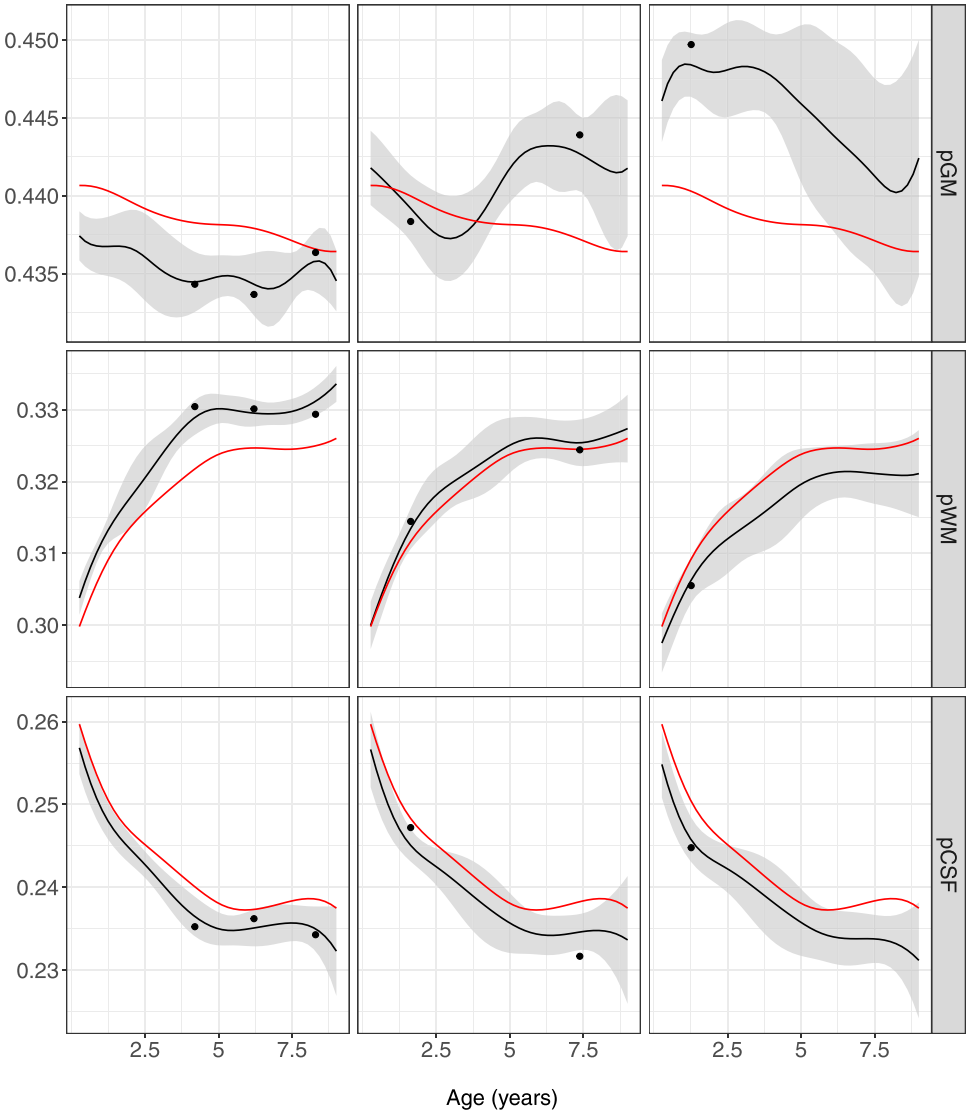

Fig. 11. Individual predicted trajectories for three randomly selected boys in the RESONANCE data. The black solid curves correspond to the fitted trajectory and grey ribbon to the 95% simultaneous confidence band around it. The red curve represents the population mean curve and solid points correspond to the observations that were used in the fitting step. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
ASMSE/10^{-5} as per (10) for the fitted trajectories by the LME and NLME models and the PACE approach, choosing the minimum numbers of FPCs that explain at least 95% FVE.

	Girls			Boys		
	pGM	pWM	pCSF	pGM	pWM	pCSF
LME model (6)	0.38	2.67	3.52	0.49	2.02	3.59
LME model (7)	0.35	1.66	2.02	0.50	1.39	2.10
NLME model (8)	0.37	1.32	2.06	0.48	0.84	1.88
PACE	**0.11**	**0.18**	**0.26**	**0.13**	**0.32**	**0.47**

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, see 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 to a participant's exclusion from the analysis.

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 (Fig. 4), eigenfunctions (Fig. 5), and modes of variation (Fig. 6) estimated using the PACE method, differentiaed by gender. As expected, raw volume measures increased in total brain volumes (TBV) as well as GM, WM and CSF with age (Figs. S.9–S.11 in the Supplement). Investigating proportional brain volumes, we demonstrated an initial pGM and pCSF decrease coupled with a pWM increase. The observed pGM decline coupled with a pWM increase reflects previous MR findings (Brain Development Cooperative Group, 2012; Giedd et al., 1999; Toga et al., 2006) and parallels cellular processes of pruning following prenatal neurogenesis, neural migration 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 develop-

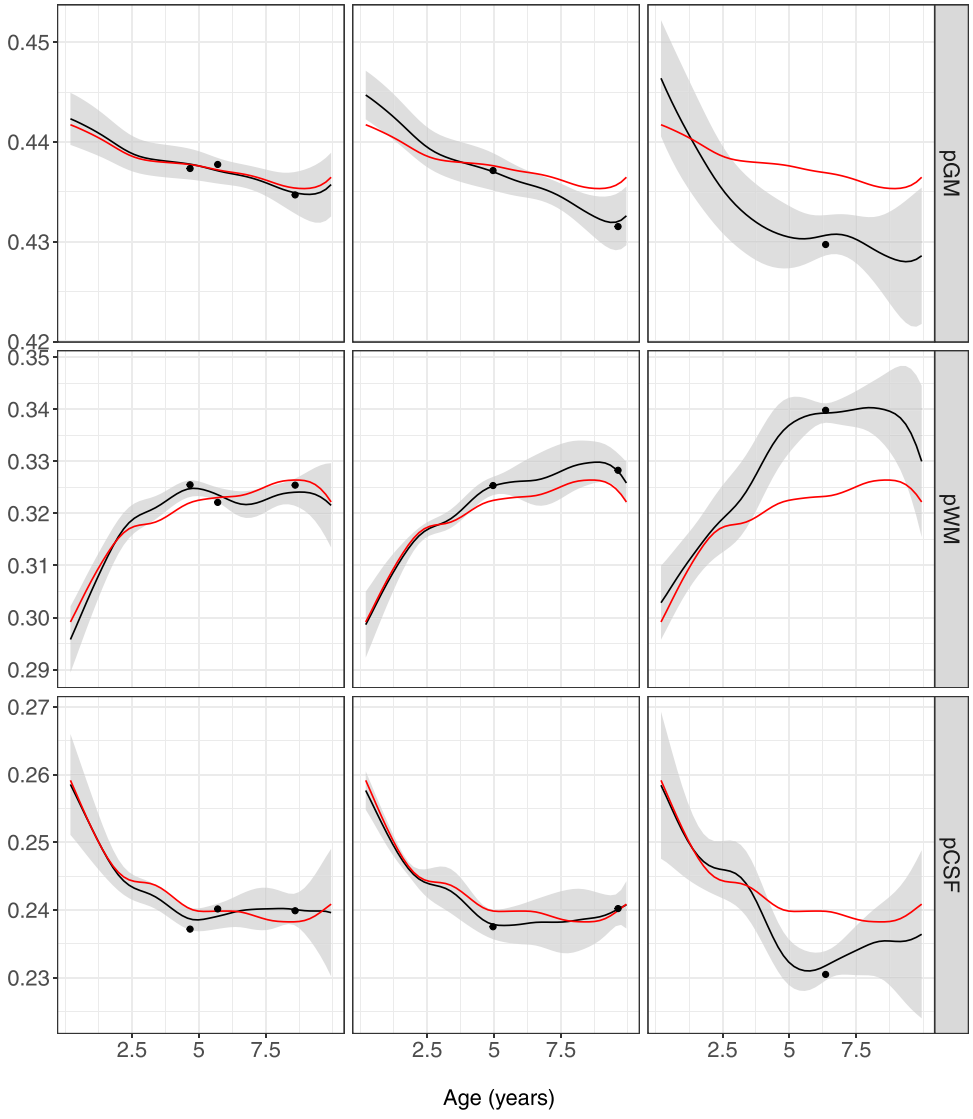

Fig. 12. Individual predicted trajectories for three randomly selected girls in the RESONANCE data. The plotting convention is same as in Fig. 11.

ment 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. We used Fréchet regression to create percentile growth charts of age-based dynamic quantiles for pGM, pWM and pCSF (Fig. 9). The resulting dynamic percentiles for boys and girls mostly evolve in similar patterns. 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 (Fig. 9). 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 (Fig. 10).

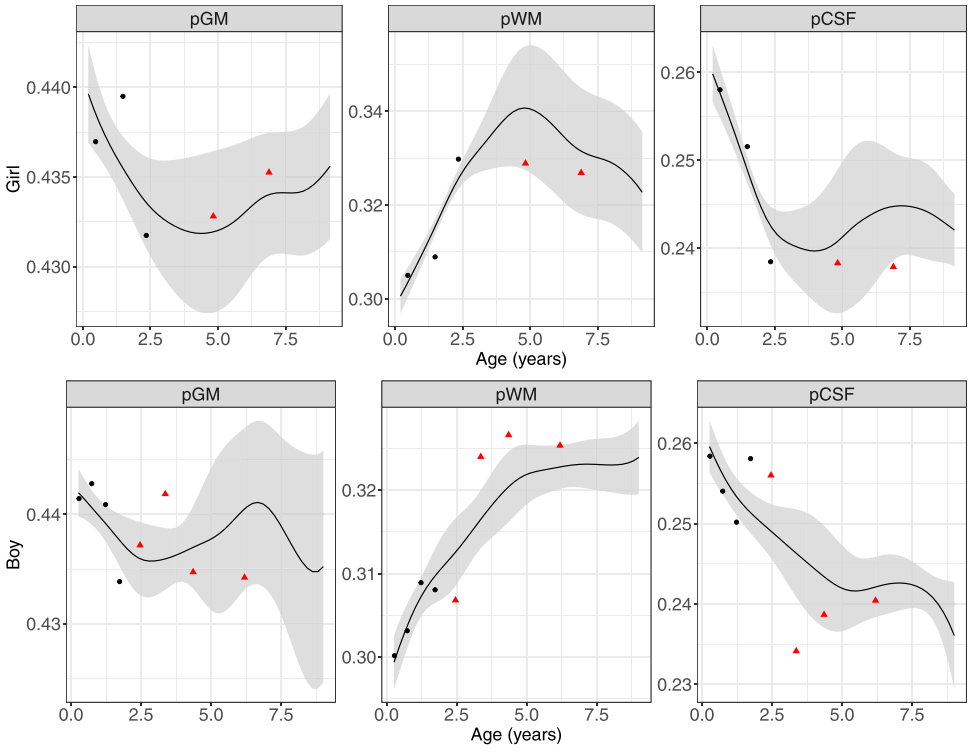

Fig. 13. The black curves correspond to the fitted trajectory and the grey ribbon to the 95% simultaneous confidence band for one selected boy (top) and one selected girl (bottom). The black round points were used in the fitting step and the red triangular points are the future measurements of the same child.

We then identified intervals where the slopes of the varying coefficient models were found to be significantly away from zero, based on pointwise bootstrap based confidence intervals, thus suggesting time zones of significant positive and negative association of brain matter proportions with cognitive development. In the younger group, higher ELC and 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 found 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 (Fig. 10). 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 levels.

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 (Figs. 11–12 respectively; Figs. S.15–S.16 in the Supplement for raw volumes). 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 reasonably close to the confidence band of the fitted trajectories (Fig. 13; Fig. S.17 in the Supplement for raw volumes). We note that although some of the future measurements of the raw volumes for the boys lie outside of the confidence bands, it is not indicative of a lack of fit. It is important to note here that the bands are for the unobservable latent trajectories, not the measurements themselves, which are contaminated by noise.

It is of interest to compare the proposed nonparametric methodology as implemented by PACE with traditional approaches for modeling and recovering individual trajectory fits for longitudinally sampled data. The most established and ubiquitous approach for the statistical analysis of data from longitudinal studies is the class of mixed effects models. Linear mixed effects (LME) models (Laird and Ware, 1982) are the most popular subclass but carry the obvious risk that they are incorrect in situations where the underlying trajectories have nonlinear shapes. This is certainly the case for the RESONANCE data where the average growth trajectories exhibit strong nonlinearity. The more general class of nonlinear mixed effects (NLME) models (Lindstrom and Bates, 1990) is handicapped by the requirement that the nonlinear function class that constitutes the fixed effects curves has to be pre-specified by the user. When the sampling design is very sparse, this is fraught with difficulties. The application of PACE completely circumvents these problems by nonparametrically recovering the mean function and the eigenfunctions, which provide an optimal function class which is then populated with random effects that correspond to the functional principal component scores. Secondly, the mean and eigenfunctions uncovered by PACE can subsequently be used to inform the construction of a nonlinear or linear random effects model for those cases where traditional users are intent to use such a parametric model class (see Section 2, Wang et al., 2016) so that even when one intends to fit a mixed effects model eventually it is advantageous to first apply PACE to determine which mixed models might be compatible with the data.

Predicting individual growth curves from the data available for individual participants data can also inform about both typical and atypical development, including early detection of early neurodevelopmental disorders or later developmental problems, which in turn can make school entry more difficult, leading to more academic struggles in the future. Especially the percentile growth curves can aid in flagging potential problems. If for example a subject's data would place this child into the 5th percentile of pWM during a time window that has been identified as sensitive for the establishment of overall cognitive func-

tioning, follow up tests and if needed early interventions could help to prevent future problems. The prediction of neurocognive development using percentiles as a population-based reference has the potential to inform 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 individual measurements being taken at random times with pairs of times from all individuals filling out the square of the domain. The success of the method depends on this assumption, which can be empirically checked by practitioners by evaluating the design plot, which is a routine output of the `fdapace` R package and is illustrated in Fig. 1 for the RESONANCE cohort. If the design plot indicates that there are major holes or snippet features appear (Lin and Wang, 2020), then covariance estimation will not be stable and neither covariance surface nor eigenfunctions can be reliably estimated. The estimated FPC scores using PACE are the best linear unbiased predictors and best predictors in the Gaussian case for the conditional expectations of the true scores, given the data observed for the subject. But they do not converge to the true scores even for large samples. This is reflected in the simultaneous bands for predicted trajectories as shown in Figs. 11 and 12. We also note that these uniform confidence bands are valid only under the assumption of Gaussianity of the true FPC scores.

Another limitation is that all findings and conclusions obtained using PACE are to some extent dependent on the bandwidth choices for smoothing of the mean functions and the covariance surfaces. Different bandwidth choices could lead to undersmoothing or oversmoothing in mean and covariance estimation, affecting the conclusions. To address this, we have developed automatic bandwidth selectors that lead to reproducible results as they do not require additional input from the user. 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.

Finally, the dynamic associations of the cognitive development scores, ELC, NVDQ and FSIQ, with pGM, pWM and pCSF do not establish causality and have not been corrected for multiple comparisons which would require the construction of simultaneous confidence bands rather than the pointwise bands illustrated in Fig. 10. The estimated concurrent effects and pointwise significance are therefore merely suggestive. Constructing theoretically valid and practically simultaneous confidence bands for the estimated slope functions is beyond the scope of the current paper and would be a topic for future research.

In addition to methodological considerations, unique challenges due to using MRI output data can arise. These can include differences in scanner make, Tesla strength and sequence as well as other factors influencing consistent measures over time. While consistent phantom measurements aimed to catch factors influencing scanning quality over time were applied following best practice procedure, the here proposed method for creating brain growth curves were used for only one cohort, thus avoiding variation in scanner make, strength and acquisition protocol. However, we are planning to use the same approach on another cohort with a different scanner make and sequence in the future, hopefully allowing us to evaluate the consistency of our findings across different cohorts.

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 for children from very sparse longitudinal data obtained in a large paediatric cohort spanning early infancy to late childhood. We also propose a method for constructing dynamic percentiles for sparsely measured brain tissue data (Fig. 9). Our methods are useful to gain insights at both the population (Figs. 4–10) and the subject (Figs. 11–13) level. While we have applied this novel analysis method to model neuroanatomical development, it can be applied more broadly to other kinds of sparse longitudinal data, including diffusion tensor imaging, functional and other magnetic resonance imaging data outputs.

Credit authorship contribution statement

Yaqing Chen: Methodology, Software, Formal analysis, Writing - original draft. **Paromita Dubey:** Methodology, Software, Formal analysis, Writing - original draft. **Hans-Georg Müller:** Conceptualization, Methodology, Software, Resources, Writing - original draft, Supervision, Funding acquisition. **Muriel Bruchhage:** Investigation, Resources, Writing - original draft. **Jane-Ling Wang:** Conceptualization, Methodology, Software, Resources, Writing - original draft, Supervision, Funding acquisition. **Sean Deoni:** Conceptualization, Formal analysis, Resources, Data curation, Supervision, Funding acquisition.

Acknowledgement

The research was supported by NIH grant UG3-0D023313 (ECHO Program) and NSF grants DMS-2014626 (H.G.M.) and DMS-1914917 (J.L.W.).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.neuroimage.2021.118079.

References

Aggarwal, O.P., 1955. Some minimax invariant procedures for estimating a cumulative distribution function. Ann. Math. Stat. 26 (3), 450–463.
Avants, B.B., Tustison, N.J., Stauffer, M., Song, G., Wu, B., Gee, J.C., 2014. The Insight ToolKit image registration framework. Front. Neuroinf. 8, 44.
Barnea-Goraly, N., Menon, V., Eckert, M., Tamm, L., Bammer, R., Karchemskiy, A., Dant, C.C., Reiss, A.L., 2005. White matter development during childhood and adolescence: a cross-sectional diffusion tensor imaging study. Cereb. Cortex 15 (12), 1848–1854.
Bates, D., Mächler, M., Bolker, B., Walker, S., 2014. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.
Bernal-Rusiel, J.L., Greve, D.N., Reuter, M., Fischl, B., Sabuncu, M.R., Alzheimer’s disease neuroimaging initiative, 2013. Statistical analysis of longitudinal neuroimage data with linear mixed effects models, 66, pp. 249–260.
Bernal-Rusiel, J.L., Reuter, M., Greve, D.N., Fischl, B., Sabuncu, M.R., Alzheimer’s disease neuroimaging initiative, 2013. Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data. Neuroimage 81, 358–370.
Blakemore, S.-J., Choudhury, S., 2006. Development of the adolescent brain: implications for executive function and social cognition. J. Child Psychol. Psychiatry 47 (3–4), 296–312.
Bondell, H.D., Reich, B.J., 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. Cereb. Cortex 22 (1), 1–12.
Bray, S., Krongold, M., Cooper, C., 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., Deoni, S.C., 2020. Functional connectivity correlates of infant and early childhood cognitive development. Brain Struct. Funct. 225 (2), 669–681.
Cai, Z., Fan, J., Li, R., 2000. Efficient estimation and inferences for varying-coefficient models. J. Am. Stat. Assoc. 95 (451), 888–902.
Carroll, C., Gajardo, A., Chen, Y., Dai, X., Fan, J., Hadjipantelis, P.Z., Han, K., Ji, H., Müller, H.-G., 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., Müller, H.-G., 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., Parzen, E., 1997. Unified estimators of smooth quantile and quantile density functions. J. Stat. Plann. Inference 59 (2), 291–307.
Cole, T., 1988. Fitting smoothed centile curves to reference data.. J. R. Stat. Soc. Ser. A 151 (3), 385–406.
Cole, T., 1994. Growth charts for both cross-sectional and longitudinal data.. Stat. Med. 13, 2477–2492.

Dai, X., Hadjipantelis, P., Wang, J.-L., Deoni, S.C., Müller, H.-G., 2019. Longitudinal associations between white matter maturation and cognitive development across early childhood. Hum. Brain Mapp. 40, 4130–4145.
Dean, D.C., Dirks, H., O'Muircheartaigh, J., Walker, L., Jerskey, B.A., Lehman, K., Han, M., Waskiewicz, N., Deoni, S.C., 2014. Pediatric neuroimaging using magnetic resonance imaging during non-sedated sleep. Pediatr. Radiol. 44 (1), 64–72.
Di Martino, A., Fair, D.A., Kelly, C., Satterthwaite, T.D., Castellanos, F.X., Thomason, M.E., Craddock, R.C., Luna, B., Leventhal, B.L., Zuo, X.-N., Milham, M.P., 2014. Unraveling the miswired connectome: a developmental perspective. Neuron 83 (6), 1335–1353.
Dong, H.-M., Castellanos, F.X., Yang, N., Zhang, Z., Zhou, Q., He, Y., Zhang, L., Xu, T., Holmes, A.J., Yeo, B.T., Chen, F., Wang, B., Beckmann, C., White, T., Sporns, O., Qiu, J., Feng, T., Chen, A., Liu, X., Chen, X., Weng, X., Milham, M.P., Zuo, X.-N., 2020. Charting brain growth in tandem with brain templates at school age. Sci. Bull. 65 (22), 1924–1934.
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. Ann. Stat. 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., Gur, R.C., 2017. Age-related effects and sex differences in gray matter density, volume, mass, and cortical thickness from childhood to young adulthood. J. Neurosci. 37 (20), 5065–5073.
Giedd, J.N., Blumenthal, J., Jeffries, N.O., Castellanos, F.X., Liu, H., Zijdenbos, A., Paus, T., Evans, A.C., Rapoport, J.L., 1999. Brain development during childhood and adolescence: a longitudinal MRI study. Nat. Neurosci. 2 (10), 861–863.
Giedd, J.N., 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., Gao, W., 2018. Imaging structural and functional brain development in early childhood. Nat. Rev. Neurosci. 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., James, A., 2010. Longitudinal changes in grey and white matter during adolescence. Neuroimage 49 (1), 94–103.
Gogtay, N., Thompson, P.M., 2010. Mapping gray matter development: implications for typical development and vulnerability to psychopathology. Brain Cognit. 72 (1), 6–15.
Gur, R.C., Calkins, M.E., Satterthwaite, T.D., Ruparel, K., Bilker, W.B., Moore, T.M., Savitt, A.P., Hakonarson, H., Gur, R.E., 2014. Neurocognitive growth charting in psychosis spectrum youths. JAMA Psychiatry 71 (4), 366–374.
Hall, P., Müller, H.-G., Wang, J.-L., 2006. Properties of principal component methods for functional and longitudinal data analysis. Ann. Stat. 34 (3), 1493–1517.
He, X., 1997. Quantile curves without crossing. Am. Stat. 51 (2), 186–192.
Hendricks, W., Koenker, R., 1992. Hierarchical spline models for conditional quantiles and the demand for electricity. J. Am. Stat. Assoc. 87 (417), 58–68.
Huang, J.Z., Wu, C.O., Zhou, L., 2004. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. Statistica Sinica 14 (3), 763–788.
Hyun, J.W., Li, Y., Huang, C., Styner, M., Lin, W., Zhu, H., Alzheimer's Disease Neuroimaging Initiative, 2016. STGP: spatio-temporal gaussian process models for longitudinal neuroimaging data. Neuroimage 134, 550–562.
Kessler, D., Angstadt, M., Sripada, C., 2016. Growth charting of brain connectivity networks and the identification of attention impairment in youth. JAMA Psychiatry 73 (5), 481–489.
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., Bassett, G., 1978. Regression quantiles. Econometrica 46, 33–50.
Koenker, R., Ng, P., Portnoy, S., 1994. Quantile smoothing splines. Biometrika 81, 673–680.
Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 963–974.
Lancaster, J.L., Tordesillas-Gutiérrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., Mazziotta, J.C., Fox, P.T., 2007. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. Hum. Brain Mapp. 28 (11), 1194–1205.
Lebel, C., Beaulieu, C., 2011. Longitudinal development of human brain wiring continues from childhood into adulthood. J. Neurosci. 31 (30), 10937–10947.
Lebel, C., Walker, L., Leemans, A., Phillips, L., 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. Ann. Inst. Stat. Math. 64 (5), 919–943.
Lin, Z., Wang, J.-L., 2020. Mean and covariance estimation for functional snippets. J. Am. Stat. Assoc. (just-accepted) 1–39.
Lindstrom, M.J., Bates, D.M., 1990. Nonlinear mixed effects models for repeated measures data. Biometrics 673–687.
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. Cereb. Cortex 28 (2), 750–763.
Matsuzawa, J., Matsui, M., Konishi, T., Noguchi, K., Gur, R.C., Bilker, W., Miyawaki, T., 2001. Age-related volumetric changes of brain gray and white matter in healthy infants and children. Cereb.l Cortex 11 (4), 335–342.
Mullen, E.M., 1995. Mullen Scales of Early Learning. Circle Pines, MN, American Guidance Service.
Müller, H.-G., Stadtmüller, U., 1999. Multivariate boundary kernels and a continuous least squares principle. J. R. Stat. Soc. Ser. B 61, 439–458.
Nobis, L., Manohar, S.G., Smith, S.M., Alfaro-Almagro, F., Jenkinson, M., Mackay, C.E., Husain, M., 2019. Hippocampal volume across age: nomograms derived from over 19,700 people in UK Biobank. Neuroimage Clin. 23, 101904.
O'Muircheartaigh, J., Dean III, D.C., Ginestet, C.E., Walker, L., Waskiewicz, N., Lehman, K., Dirks, H., Piryatinsky, I., Deoni, S.C., 2014. White matter development and early cognition in babies and toddlers. Hum. Brain Mapp. 35 (9), 4475–4487.
Panaretos, V.M., Zemel, Y., 2016. Amplitude and phase variation of point processes. Ann. Stat. 44 (2), 771–812.
Parzen, E., 1979. Nonparametric statistical data modeling. J. Am. Stat. Assoc. 74 (365), 105–121.
Petersen, A., Müller, H.-G., 2016. Functional data analysis for density functions by transformation to a Hilbert space. Ann. Stat. 44 (1), 183–218.
Petersen, A., Müller, H.-G., 2019. Fréchet regression for random objects with Euclidean predictors. Ann. Stat. 47 (2), 691–719.
Peterson, M., Warf, B.C., Schiff, S.J., 2018. Normative human brain volume growth. J. Neurosurg. Pediatr. 21 (5), 478–485.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2020. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-148.
Pinheiro, J.C., Bates, D.M., 2020. Linear mixed-effects models: basic concepts and examples. In: Mixed-Effects Models in S and S-PLUS. Springer, pp. 3–56.
Read, R., 1972. The asymptotic inadmissibility of the sample distribution function. Ann. Math. Stat. 43 (1), 89–95.
Rigby, R.A., Stasinopoulos, D.M., 2004. Smooth centile curves for skew and kurtotic data modelled using the Box–Cox power exponential distribution. Stat. Med. 23 (19), 3053–3076.
Sadeghi, N., Prastawa, M., Fletcher, P.T., Vachet, C., Wang, B., Gilmore, J., Gerig, G., 2013. Multivariate modeling of longitudinal MRI in early brain development with confidence measures. In: 2013 IEEE 10th International Symposium on Biomedical Imaging. IEEE, pp. 1400–1403.
Samanta, M., 1989. Non-parametric estimation of conditional quantiles. Stat. Probab. Lett. 7 (5), 407–412.
Sanford, R., Ances, B.M., Meyerhoff, D.J., Price, R.W., Fuchs, D., Zetterberg, H., Spudich, S., Collins, D.L., 2018. Longitudinal trajectories of brain volume and cortical thickness in treated and untreated primary human immunodeficiency virus infection. Clin. Infect. Dis. 67 (11), 1697–1704.
Şentürk, D., Müller, H.-G., 2008. Generalized varying coefficient models for longitudinal data. Biometrika 95 (3), 653–666.
Şentürk, D., Müller, H.-G., 2010. Functional varying coefficient models for longitudinal data. J. Am. Stat. Assoc. 105 (491), 1256–1264.
Şentürk, D., 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., 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 Anal. Manifolds GraphsMetric Spaces (Paris, 2002) 338, 357–390.
Tamnes, C.K., Herting, M.M., Goddings, A.-L., Meuwese, R., Blakemore, S.-J., Dahl, R.E., Güroğlu, B., Raznahan, A., Sowell, E.R., Crone, E.A., et al., 2017. Development of the cerebral cortex across adolescence: a multisample study of inter-related longitudinal changes in cortical volume, surface area, and thickness. J. Neurosci. 37 (12), 3402–3412.
Toga, A.W., Thompson, P.M., Sowell, E.R., 2006. Mapping brain maturation. Focus 29 (3), 148–390.
UNICEF, WHO, 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., Kremen, W.S., 2018. Brain structure mediates the association between height and cognitive ability. Brain Struct. Funct. 223 (7), 3487–3494.
Wang, J.-L., Chiou, J.-M., Müller, H.-G., 2016. Functional data analysis. Annu. Rev. Stat. Appl. 3, 257–295.
Wechsler, D., 2012. Wechsler Preschool and Primary Scale of Intelligence', fourth ed. 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., 2010. Varying-coefficient functional linear regression. Bernoulli 16 (3), 730–758.
Yang, S.-S., 1985. A smooth nonparametric estimator of a quantile function. J. Am. Stat. Assoc. 80 (392), 1004–1011.
Yao, F., Müller, H.-G., Wang, J.-L., 2005. Functional data analysis for sparse longitudinal data. J. Am. Stat. Assoc. 100 (470), 577–590.
Yu, Q., Peng, Y., Kang, H., Peng, Q., Ouyang, M., Slinger, M., Hu, D., Shou, H., Fang, F., Huang, H., 2020. Differential white matter maturation from birth to 8 years of age. Cereb. Cortex 30 (4), 2674–2690.