FRÉCHET ESTIMATION OF TIME-VARYING COVARIANCE MATRICES FROM SPARSE DATA, WITH APPLICATION TO THE REGIONAL CO-EVOLUTION OF MYELINATION IN THE DEVELOPING BRAIN

BY ALEXANDER PETERSEN, SEAN DEONI[1] AND HANS-GEORG MÜLLER[2]

University of California, Santa Barbara, Brown University and University of California, Davis

Assessing brain development for small infants is important for determining how the human brain grows during the early period of life when the rate of brain growth is at its peak. The development of MRI techniques has enabled the quantification of brain development. A key quantity that can be extracted from MRI measurements is the level of myelination, where myelin acts as an insulator around nerve fibers and its deployment makes nerve pulse propagation more efficient. The co-variation of myelin deployment across different brain regions provides insights into the co-development of brain regions and can be assessed as correlation matrix that varies with age. Typically, available data for each child are very sparse, due to the cost and logistic difficulties of arranging MRI brain scans for infants. We showcase here a method where data per subject are limited to measurements taken at only one random age, so that one has cross-sectional data available, while aiming at the time-varying dynamics. This situation is encountered more generally in cross-sectional studies where one observes p-dimensional vectors at one random time point per subject and is interested in the $p \times p$ correlation matrix function over the time domain. The challenge is that at each observation time one observes only a p-vector of measurements but not a covariance or correlation matrix. For such very sparse data, we develop a Fréchet estimation method. Given a metric on the space of covariance matrices, the proposed method generates a matrix function where at each time the matrix is a nonnegative definite covariance matrix, for which we demonstrate consistency properties. We discuss how this approach can be applied to myelin data in the developing brain and what insights can be gained.

1. Introduction. Modern Magnetic Resonance Imaging (MRI) methodology has made it possible to study structural elements of the brain for small infants. Of special interest is to utilize this technique for the study of how the brain grows and

develops in the early childhood years, a period of rapid brain and cognitive growth [Lenroot and Giedd (2006)], during which the brain's structural and functional networks are formed. The maturity of the developing brain is customarily measured in terms of level of myelination, which can be extracted from MRI signals. Myelin serves as an insulating sheath around nerve fibers and higher myelin levels are thought to be associated with more efficient signaling and thus improved brain function. Altered brain connectivity has been hypothesized to underlie a number of intellectual, behavioral, and psychiatric disorders [Stiles and Jernigan (2010), Lebel and Beaulieu (2011), Shaw et al. (2008)]. To evaluate the development of brain connectivity, the evolution of correlations between myelin levels for various white matter brain regions as a function of age in normal children is of interest. This is because brain development can be characterized by the degree of myelination and time-varying correlation matrices that quantify the dynamic correlations of myelination between different brain regions pinpoint the regions that are developing at similar age periods, characterizing the spatio-temporal brain development map.

A general problem for the analysis of MRI studies in children is the sparsity of data in the time domain, which is a consequence of cost and logistic difficulties. Even in studies that were intended to be longitudinal, for many infants only measurements at one point in time are available. This motivates the development of the approach that we present here, where we obtain consistent estimates of the underlying time-varying correlation matrices from sparse observations, where the estimates are guaranteed to be symmetric and nonnegative definite matrices at each time argument. We consider here the sparsest possible case, namely that each subject is observed at only one random age. Our methodology is of interest for any study where one aims at time-dynamic information for the correlation/covariance structure of continuous multivariate observations, when one essentially has only one measurement per subject available.

In our application to brain development, the underlying process of each subject corresponds to the myelination levels for $p = 21$ brain regions, which evolve continuously over time. However, observations are measured only at a single time point, so that the data available for each subject are the (random) time point at which the measurement was taken as well as a multivariate p-vector of region-specific myelin levels. As the data are vectors from which the smoothly varying correlation matrix is the desired target, we show here that an adaptation of Fréchet regression [Petersen and Müller (2018)] provides a simple and effective approach to recover the correlation matrices. The proposed method produces for each age an estimate of the population correlation matrix with the following properties: the estimates are correlation matrices themselves; and they are consistent, with convergence rates that can be derived from the general theory of Fréchet regression.

In Section 3, we will demonstrate these methods with data obtained as part of an on-going accelerated longitudinal neuroimaging study, the Brown University Assessment of Myelination and Behaviour Across Maturation (BAMBAM) study

[Deoni et al. (2011, 2015)], by estimating time-varying correlation matrix functions. The cross-correlation matrices of myelination level in dependency on age entail a spatio-temporal map of development that indicates the locations of early and late development, as well as the changing connectivity of various sets of regions of interest. For the children in the BAMBAM study, at each age where a MRI was obtained, the Early Learning Composite (ELC) score is also available, which quantifies neurocognitive development. As an application of the estimated covariance structure between the myelination levels of various brain regions across age, we use this estimated covariance matrix, in addition to a similarly constructed cross-correlation between the ELC outcome and the brain region myelin levels, to obtain linear regressions of ELC versus the actual myelination levels for the $p = 21$ brain regions. Since the coefficients in the linear regression are also time-dependent, without the age-dependent covariance and cross-covariance estimates this regression would not be possible as for each age where predictors (brain myelination levels) and response (ELC) are available, we have only one measurement. Our analysis pinpoints the relevance of various brain regions for neurocognitive outcomes in dependence of the age of infants.

Previous work on modeling and quantifying brain development has not addressed the modeling of cross-correlations and covariances, with the exception of Dai et al. (2017a), where ad hoc kernel smoothing techniques were employed to obtain pairwise cross-correlation functions for the sparsely observed BAMBAM data. However, the ensemble of the recovered pairwise correlations does not form a valid covariance/correlation matrix function, which at each age argument is required to be symmetric and nonnegative definite [Petersen and Müller (2016)]. Of these properties, symmetry can be easily guaranteed, while nonnegative definiteness is a different matter and is not guaranteed for pairwise correlation function estimation approaches.

In this paper, we study kernel smoothing methods for the coherent estimation of these time-varying covariance matrices, where "coherent" means that the estimators are symmetric nonnegative definite matrices at all time points. In Section 2.2, we describe existing smoothing techniques which allow one to estimate the individual elements of each covariance matrix and discuss their drawbacks in the current setting. Sections 2.3 and 2.4 reformulate the problem as the smoothing of entire covariance matrices, opening a connection to a recent method for regression for complex objects, the *Fréchet regression* model. Section 3 gives more detailed background on the BAMBAM study and the Fréchet estimation technique is applied to study dynamic myelin correlations. We also demonstrate that the estimated covariance matrix functions are instrumental to establish the relation between the myelination levels over the 21 brain regions with concurrent ELC scores. Section 4 validates the approach by demonstrating in a simulation study the ability of the Fréchet estimator to consistently recover dynamic covariances from such sparse data using similar sample size and dimension, and its superiority to an alternative kernel estimator. Section 5 demonstrates theoretical advantages of the Fréchet

smoother and extends theory for consistent estimation, which, by necessity, must be performed using biased and degenerate covariance matrices as only multivariate random vectors are observed.

2. Regression with covariance as response.

2.1. *Preliminaries.* We begin by defining the data objects and targets for estimation. The observed data consist of i.i.d. pairs (X_i, Y_i), $i = 1, \ldots, n$, where X_i represents the age at which the child was examined and $Y_i \in \mathbb{R}^p$ is a vector that contains the myelin water fraction (MWF) levels at $p = 21$ brain regions measured during the visit (only available at one age). It is natural to think of Y_i as a point observation of a latent smooth multivariate stochastic process $\{U_i(x) \in \mathbb{R}^p : x \in [0, T]\}$, where the U_i are i.i.d. mean-square continuous processes [Ash and Gardner (1975), Hsing and Eubank (2015)], and $Y_i = U_i(X_i)$. Thus, the process U_i tracks the MWF levels of subject i at all ages, and Y_i is the value of this process at a random age X_i that is independent of the process U_i. The targets of interest are the cross-sectional mean and covariance of the latent process, which can be expressed in terms of the conditional mean and covariance of $Y_i | X_i$ via

$$
\begin{aligned}
\mu(x) &= E(Y_i | X_i = x) = E\big(U_i(x)\big), \\
\Sigma(x) &= \operatorname{Cov}(Y_i | X_i = x) = \operatorname{Cov}\big(U_i(x)\big).
\end{aligned}
\tag{2.1}
$$

Notice that the covariance target is a function $\Sigma : [0, T] \to \mathcal{S}_p$, where $\mathcal{S}_p$ is the space of symmetric nonnegative definite matrices of dimension p. Similarly, one is interested in the corresponding pointwise correlation matrices $\mathcal{R}(x)$. Estimators $\hat{\mu}_j(x)$ of the component functions $\mu_j(x)$ of $\mu(x)$ can be obtained in a variety of ways, including smoothing splines and local polynomial methods or other scatterplot smoothers, and have been well explored while this is not the case for the estimation of $\Sigma(x)$. Therefore we will focus our attention on the covariance estimation.

2.2. *Cross-covariance estimation.* Established nonparametric techniques could be invoked for the estimation of the individual matrix elements $\Sigma_{jk}(x)$ by smoothing so-called *raw covariances*

$$
c_{ijk} = \big(Y_{ij} - \hat{\mu}_j(X_i)\big)\big(Y_{ik} - \hat{\mu}_k(X_i)\big).
\tag{2.2}
$$

Let K be a univariate density function, h_{jk} a bandwidth and $K_h(y) = h^{-1} \times K(yh^{-1})$. Local constant (Nadaraya–Watson) and local linear smoothing are two common tools that can be enlisted to smooth these raw covariances, with respective estimators $\hat{\Sigma}^0_{jk}(x) = \hat{a}$ and $\hat{\Sigma}^1_{jk}(x) = \hat{b}$, where

$$
\hat{a} = \operatorname*{argmin}_{a \in \mathbb{R}} \sum_{i=1}^{n} K_{h_{jk}}(X_i - x)(c_{ijk} - a)^2,
$$

$$(\hat{b}, \hat{c}) = \underset{b,c \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} K_{h_{jk}}(X_i - x)\big(c_{ijk} - b - c(X_i - x)\big)^2.$$

As solutions to weighted least squares problems, both of these kernel estimators take the form of a weighted average [Fan and Gijbels (1996)],

$$(2.3) \qquad \hat{\Sigma}_{jk}^m(x) = \sum_{i=1}^{n} w_{in,m} c_{ijk},$$

where $m = 0, 1$ correspond to Nadaraya–Watson and local linear smoothing, respectively. Letting

$$r_\ell = \frac{1}{n} \sum_{i=1}^{n} K_{h_{jk}}(X_i - x)(X_i - x)^\ell, \qquad \ell = 0, 1, 2,$$

and $\sigma^2 = r_0 r_2 - r_1^2$, the weights $w_{in,m}$ in (2.3) are, respectively,

$$(2.4) \qquad w_{in,0} = w_{in,0}(x, h_{jk}) = \frac{K_{h_{jk}}(X_i - x)}{\sum_{i=1}^{n} K_{h_{jk}}(X_i - x)} \quad \text{and}$$

$$(2.5) \qquad w_{in,1} = w_{in,1}(x, h_{jk}) = \frac{1}{\sigma^2} K_{h_{jk}}(X_i - x)\big[r_2 - r_1(X_i - x)\big],$$

both satisfying $\sum_{i=1}^{n} w_{in,m} = 1$. This approach is closely related to the method used in Dai et al. (2017a).

The above described estimators are known to be consistent on the interior $(0, T)$, while the local linear estimator is preferable due to its improved performance near the boundaries, which can make a substantial difference in practical applications [Fan and Gijbels (1996)]. However, in our application, these estimators possess some undesirable properties when combined into matrix estimators $\hat{\Sigma}^m(x)$, $m = 0, 1$. First, while both lead to symmetric matrices, neither is guaranteed to be nonnegative definite, so that standard analyses involving covariance matrices cannot be applied without some post-hoc alterations.

In the case of the Nadaraya–Watson estimator, this can in fact be resolved by setting a common bandwidth $h_{jk} = h$. Since the weights $w_{in,0}$ are strictly positive and the space of covariance matrices is convex, this will give a coherent unified estimator $\hat{\Sigma}^m(x) \in \mathcal{S}_p$. However, this apparent fix does nothing to resolve poor performance near the boundaries. On the other hand, even with a common bandwidth, the local linear estimator can still fail to be nonnegative definite, since $w_{in,1}$ can be (and often are) negative, especially near the boundaries. If one requires a true covariance, a natural procedure would be to project onto the space of covariance matrices by truncating negative eigenvalues to zero. In the next section, we justify this approach by reframing the covariance estimation problem as a geometric problem involving Fréchet means and an adapted version of a recently developed regression method for responses that are random objects in a metric space.

2.3. *Time-varying covariance matrices as conditional Fréchet means.* A key step is to characterize the matrix $\Sigma(x)$ as a conditional Fréchet mean. The Fréchet mean of a random element Z of an arbitrary metric space (Ω, d) is

$$(2.6) \qquad E_\oplus(Z) = \operatorname*{argmin}_{\omega \in \Omega} E\big(d^2(Z, \omega)\big),$$

where existence can be guaranteed by compactness, but uniqueness cannot. If Ω is a convex subset of a Euclidean space, and d is the Euclidean metric, then the ordinary mean and Fréchet mean are known to coincide. Let U be a random p-vector with zero mean and covariance matrix Λ and set $Z = UU^\top$. Then, if $\Omega = \mathcal{S}_p$ and $d = d_F$ is the Frobenius distance, it can be easily shown that the unique Fréchet mean of Z is $E_\oplus(Z) = \Lambda$ due to the fact that d_F is a Euclidean norm as it arises from an inner product.

Now, let X and U be generic copies of the sample elements X_i and the latent multivariate processes U_i, $Y = U(X)$ and $C(X) = (Y - \mu(X))(Y - \mu(X))^\top$. Then, for the Frobenius metric d_F, the matrix $\Sigma(x)$ in (2.1) can naturally be expressed as

$$(2.7) \qquad \Sigma(x) = E\big(C(X)|X = x\big) = E_\oplus\big(C(X)|X = x\big).$$

This characterization suggests that, while smoothing is an appropriate method for estimation of $\Sigma(x)$, it should not be done on the scalar raw covariances in (2.2), as this merely targets the individual entries of $\Sigma(x)$. Rather, one should smooth entire covariance matrices and target the Fréchet mean $\Sigma(x)$. As our data consist of multivariate observations, the data objects closest to our targets are the raw covariance matrices

$$(2.8) \qquad \hat{C}_i = \big(Y_i - \hat{\mu}(X_i)\big)\big(Y_i - \hat{\mu}(X_i)\big)^\top.$$

Before we develop the estimator, we provide two remarks that highlight the unique difficulties and challenges associated with this estimation problem.

REMARK 1. The covariances $\hat{C}_i$ are *biased* in the sense that

$$E_\oplus(\hat{C}_i|X_i = x) \neq \Sigma(x).$$

This is because they serve only as approximations for the unobservable quantities

$$(2.9) \qquad C_i = \big(Y_i - \mu(X_i)\big)\big(Y_i - \mu(X_i)\big)^\top,$$

for which $E_\oplus(C_i|X_i = x) = \Sigma(x)$.

REMARK 2. Another feature of the raw covariances $\hat{C}_i$, which we would still face if we actually observed the unknown C_i in (2.9), is that they are degenerate, that is, they have rank one. The challenge is then that we are attempting to estimate the object $\Sigma(x)$ that lies in the interior of $\mathcal{S}_p$ using a sample of objects that, by definition, reside on the boundary of the space.

In recent years, the smoothing of covariance matrices has been studied in various contexts, with two prominent examples being the mapping of neural pathways using DTI [Yuan et al. (2012), Carmichael et al. (2013)] and the spatial modeling of speech recordings [Tavakoli et al. (2016)]. In the DTI application, the data consist of a random sample of pairs (P_i, S_i), where $P_i \in \mathcal{R}$ and S_i is a symmetric positive definite matrix, and the target is the conditional Fréchet mean $E_\oplus(S_i | P_i = p)$ under two particular Riemannian metrics for $\mathcal{S}_p$. Being based on the sampling of nondegenerate covariance matrices, due to the availability of much richer data in this context, these previous approaches were not affected by the challenges summarized in Remarks 1 and 2. Furthermore, since Fréchet means generally depend on the chosen metric, in these applications specific features of the chosen metric could be exploited. This is in contrast to our situation, where the goal is to gain an understanding of the time-varying covariability of multivariate data. The canonical metric for this purpose is the Frobenius or L^2 metric for covariance matrices.

The importance of estimating time-varying covariance matrices is also evident in the spatial analysis of sound objects in Tavakoli et al. (2016), where the covariance also has a spatial component. Here, the smoothing of covariances takes place only over space and, at least conceptually, the objects that are smoothed are not random but fixed. In practice, these objects need to be estimated, and assumptions are included under which the estimation error is negligible. Such assumptions are common in functional data settings, but do not apply to the cross-sectional data that we consider here, where the raw data are decidedly inconsistent. In short, the nature of our data requires overcoming the challenges that are highlighted in the above remarks, whereas these challenges do not arise in covariance estimation settings that have been previously studied.

Tavakoli et al. (2016) also introduced the so-called d-covariance for multivariate data. In our setting, if d is a metric on $\mathcal{S}_p$, the d-covariance is

$$(2.10) \qquad \Sigma_\oplus^d(x) = \underset{S \in \mathcal{S}_p}{\operatorname{argmin}} E\big(d^2(C(X), S) | X = x\big).$$

The motivation for the d-covariance is that other metrics besides the Frobenius metric are more suitable for analysis on $\mathcal{S}_p$, which is a nonlinear Riemannian manifold. As Tavakoli et al. (2016) demonstrated, if d is the square-root metric $d(C_1, C_2) = d_F(C_1^{1/2}, C_2^{1/2})$, one can readily use linear methods on the square-root space, which can be identified with the linear space of symmetric matrices. The d-covariance thus provides a natural class of covariance objects that can be estimated using Fréchet methods. Alternative metrics that would have similar limitations as the d-covariance are those in the Box–Cox class, where $d_\alpha(C_1, C_2) = d_F(C_1^\alpha, C_2^\alpha)$, with appropriately defined matrix logarithms for the case $\alpha = 0$ [Pigoli et al. (2014), Petersen and Müller (2016)].

However, since $\Sigma(x) = \Sigma_\oplus^d(x)$ if and only if d is the Frobenius metric, one cannot smooth under an alternative metric without losing the ability to interpret the target as a true covariance of the observed random vector. In the particular

application of these methods to the study of myelination that is our primary motivation, we work with the Frobenius metric for two reasons. First, practitioners are more comfortable with assessing ordinary covariance and correlation between myelin levels due to its well-established interpretation. Second, many statistical models that one might wish to apply to the study of myelin levels necessitate an estimate of their covariance properties, which would not be provided when applying a smoothing procedure under a different metric, as this results in estimates that target a different quantity. The motivating problem of regressing ELC scores, which assess cognitive development, on myelin levels as predictors provides one such instance.

While we consider Fréchet methods under the Euclidean or Frobenius metric, it bears emphasizing that our approach is not restricted to this choice. Indeed, the theory in Section 5 is based on results in Petersen and Müller (2018), which hold under very general conditions and can readily be extended to the Fréchet estimation of d-covariances for various metrics d.

2.4. *Fréchet estimation.* The expression of $\Sigma(x)$ as a conditional Fréchet mean in (2.7) suggests estimation by a locally weighted sample Fréchet mean [see, e.g., Pigoli et al. (2014)]

$$(2.11) \qquad \hat{\Sigma}^{\text{NW}}(x) = \underset{C \in \mathcal{S}_p}{\operatorname{argmin}} \sum_{i=1}^{n} K_h(X_i - x) d_F^2(C, \hat{C}_i).$$

The individual components of this estimator coincide with those in (2.3) for $m = 0$, with weights as in (2.4) and all bandwidths h_{jk} being equal. So, while the standard Nadaraya–Watson smoother does not incorporate the constraints of the space $\mathcal{S}_p$, it results in a standard Fréchet estimate of $\Sigma(x)$, due to convexity. As mentioned before, this estimator does not work well near boundaries due to bias, and a local linear type estimator based on weights (2.5) is preferable.

Local linear kernel estimation does not immediately generalize to nonlinear spaces, regardless of whether they are convex or not, as it introduces negative weights near the boundaries, which is necessary to control bias. In Yuan et al. (2012), intrinsic local polynomial regression (ILPR) was proposed using manifold features of $\mathcal{S}_p$ under metrics different from d_F. Since the Frobenius metric is the natural choice given our target, ILPR is not directly applicable and, more importantly, more complex than necessary. A simpler, yet more general, re-characterization of local linear smoothing for arbitrary metric spaces (Local Fréchet Regression or Fréchet smoothing) has recently been developed in Petersen and Müller (2018). Fréchet smoothing can be easily applied to the current problem of smoothing covariance matrices, where it becomes

$$(2.12) \qquad \hat{\Sigma}^{\text{LF}}(x) = \underset{C \in \mathcal{S}_p}{\operatorname{argmin}} \sum_{i=1}^{n} w_{in,1}(x, h) d_F^2(C, \hat{C}_i),$$

and is represented as a weighted Fréchet mean, with weights $w_{in,1}$ as in (2.5).

In Fréchet smoothing, we are not weighting the covariances directly, but rather the squared distances which define the Fréchet functional to be minimized. Again, these weights can be negative, since they are derived from the local linear weights in a standard Euclidean smoothing problem, so that $\sum_{i=1}^{n} w_{in,1} \hat{C}_i$ is not guaranteed to be in $\mathcal{S}_p$. However, similar to the Nadaraya–Watson estimator, the local Fréchet estimator has an analytic expression in the case of covariance matrices and the Frobenius metric, given by

$$\hat{\Sigma}^{\mathrm{LF}}(x) = \Pi_{\mathcal{S}_p} \left(\sum_{i=1}^{n} w_{in,1}(x, h) \hat{C}_i \right),$$

where $\Pi_{\mathcal{S}_p}$ is the projection operator onto $\mathcal{S}_p$ under d_F. This projection operator is easy to compute, as it corresponds to truncating negative eigenvalues to zero while keeping the eigenspaces the same. It is interesting, albeit unsurprising, that, while $\hat{\Sigma}^{\mathrm{LF}}(x)$ as defined in (2.12) is clearly an intrinsic estimator as a minimizer of a Fréchet functional, it is also equivalent to the projection onto $\mathcal{S}_p$ of the ordinary, extrinsic local linear estimator. This phenomenon is explained by the fact that Fréchet means under Euclidean metrics coincide with ordinary means, along with the uniqueness of Euclidean projections onto closed, convex spaces such as $\mathcal{S}_p$.

3. Regional co-evolution of brain myelination in the developing brain.

3.1. *Background and the BAMBAM study.* Altered early brain development is hypothesized to underlie many neurological, behavioral, and intellectual disorders. Infancy and early childhood are sensitive periods of brain growth, coinciding with the emergence of nearly all cognitive, behavioral, and social functioning abilities. Beginning in utero, myelination advances rapidly over the first 2 years of life in a carefully choreographed caudal-cranial, posterior-to-anterior pattern [Brody et al. (1987), Paus et al. (2001)], and continues throughout childhood and adolescence [Bartzokis et al. (2010)]. This maturation pattern is driven by the tight regulation of myelination by neural activity [Fields (2005), Ishibashi et al. (2006), Lang et al. (2003)], which likely underpins its spatio-temporal coincidence with emerging neurobehavior [Fields (2008), Nagy, Westerberg and Klingberg (2004), Casey, Galvan and Hare (2005)].

The prolonged nature of myelination imparts a high degree of flexibility and plasticity to developing neural systems. However, this protracted timeline of development also places these systems at prolonged risk to injury or deviant development [Rodier (1995)]. It is increasingly recognized that alterations in myelination timing and/or extent can significantly affect behavioral and cognitive outcomes [Fields (2008)], with altered white matter microstructure being a consistent finding in many neurological and neuropsychiatric disorders [Wolff et al. (2012), Flynn et al. (2003), Xiao et al. (2014)]. Despite the importance of coordinated neural communication, relatively little is known regarding the development of structural and functional networks in infants or toddlers or the relationships linking the emergence of neural networks to evolving cognition or neurobehavioral outcomes.

The Brown University Assessment of Myelination and Behavior Across Maturation (BAMBAM) study provides voxelwise data on the myelination for neurotypical children that were very sparsely measured across age, and which we analyze in Section 3.2 to demonstrate some of these issues. Only children with measurements before the age of 1250 days were included, resulting in a sample size of $n = 223$ children. Additionally, for those children who were assessed at multiple time points during this period, one observation per subject was selected randomly.

Imaging measures include anatomical T_1-weighted (T_1w), quantitative relaxometry (qT_1, qT_2), and quantitative myelin water fraction (MWF), among other measurements. The children included in the study were devoid of major risk factors for neurologic and psychiatric disorders. Children with in utero alcohol or illicit substance exposure, premature (<37 weeks gestation) or multiple birth, fetal ultrasound abnormalities, complicated pregnancy, APGAR scores <8, NICU admission, neurological disorder, psychiatric or developmental disorders in the infant, parents or siblings were excluded [Chlebowski et al. (2013), Bilenberg (1999)].

The BAMBAM study acquired high-quality and artifact-free MRI data that inform on different characteristics of the developing brain [Fornari et al. (2007), Miller et al. (2012), Hagmann et al. (2010), Grydeland et al. (2013), Shafee, Buckner and Fischl (2015)], including morphometric (whole-brain, white and gray matter volume), myelin water fraction (which informs more specifically on myelination, a key process of brain connectivity), and other modalities [Deoni et al. (2008a), Deoni, Rutt and Peters (2006), Deoni et al. (2015)]. Such high quality MRI measurements (see Figure 1) are a prerequisite to assess brain development [Knickmeyer et al. (2008), Deoni et al. (2008b), Deoni et al. (2011), Deoni, Rutt and Peters (2003), Deoni (2007), Deoni et al. (2004), Deoni (2011), Deoni, Peters and Rutt (2004), Deoni and Kolind (2015)]. It is also of interest to relate myelin
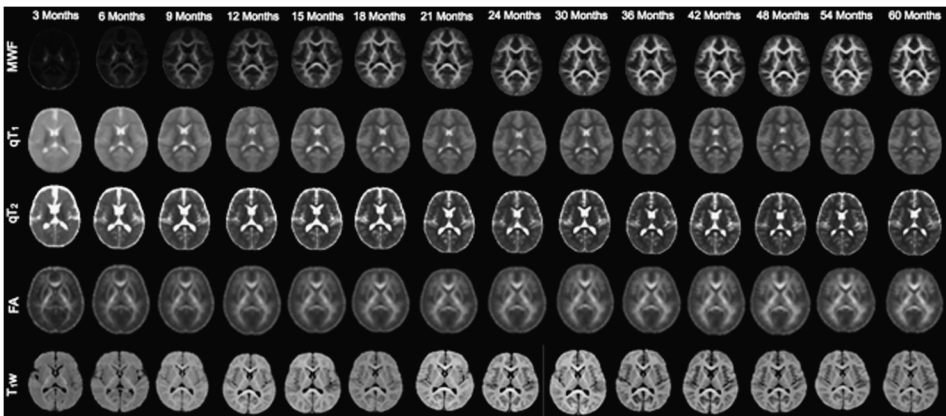

FIG. 1. *A multi-modal assessment of development. Mean myelin water fraction (*MWF*), quantitative relaxometry (*qT_1*, *qT_2*) and fractional anisotropy (*FA*) maps, and T_1-weighted (*T_1w*) images from three months to five years of age.*

level correlation across brain regions to concurrent developmental and cognitive scores [Shaw et al. (2006)], where we use the Early Learning Composite (ELC), derived from the verbal and nonverbal development quotient (VDQ and NVDQ, respectively) scores of the Mullen Scales of Early Learning [Mullen et al. (1995)]; these scores are standardized for age. It is of special interest to identify the brain regions for which myelination levels are most predictive of these scores at certain ages. This will help to pinpoint the relative relevance of brain regions for neurocognitive performance across age, a goal that has been elusive so far.

The implementation of Fréchet regression that we propose here is a one-step approach that leads directly to bona fide correlation/covariance matrix functions and is supported by theoretical consistency and rate of convergence results that are applicable even for the extremely sparse data case that we consider here. We demonstrate that also in practice this method leads to sensible results when applied to the very sparse data case, where the observed data do not correspond to a sample covariance matrix paired with each predictor, but rather just one random vector of observed levels for each predictor level. Our goal is to quantify the evolution of the cross-correlations with age from data that contain only one measurement per subject. For this, we select $p = 21$ discrete anatomical brain regions as an example. The methods generalize to other local aggregations of the MWF levels, which are recorded at the voxel level. These regions are listed in Table 1 together with their acronyms used throughout the remainder of the paper.

While modeling cross-correlation and cross-covariance of myelination has not been well explored, a related topic with more substantial work is the modeling of the myelin trajectories as a function of age. Nonlinear approaches have been applied previously to model overall myelination level trajectories by averaging over brain regions and focusing on the dependence of myelin levels as a function of

TABLE 1
List of all $p = 21$ regions/tracts and their acronyms for which MWF levels are measured

Region	Acronym	Region	Acronym
body Corpus Callosum	bCC	left Cingulum	lC
genu Corpus Callosum	gCC	right Cingulum	rC
splenu Coropus Collosum	sCC	left Corona Radiata	lCR
left Frontal	lF	right Corona Radiata	rCR
right Frontal	rF	left Internal Capsule	lIC
left Parietal	lP	right Internal Capsule	rIC
right Parietal	rP	left Optic Radiation	lOR
left Occipital	lO	right Optic Radiation	rOR
right Occipital	rO	left Superior Longitudinal Fasciculus	lSLF
left Temporal	lT	right Superior Longitudinal Fasciculus	rSLF
right Temporal	rT		

age [Shaw et al. (2008), Knickmeyer et al. (2008), Casey, Galvan and Hare (2005), Deoni et al. (2016), Deoni et al. (2015)]. Recently, Dai et al. (2017b) proposed a method to obtain whole brain MWF trajectories, modifying approaches of Şentürk and Müller (2010), Şentürk and Nguyen (2011). For the case where repeated measurements are available for each subject, ideally densely spaced, related work on modeling multivariate trajectories by functional data analysis methods includes Zhou, Huang and Carroll (2008), Chiou and Müller (2016). However, these approaches are not applicable if only one or very few observations per subject are available.

3.2. *Age-varying correlations of regional myelination.* The BAMBAM study cohort includes $n = 233$ children who were observed during the first 1250 days of life, with the earliest observation at age 65 days. Some children were observed repeatedly within this time period for a total of $N = 440$ visits, although nearly half (48%) of the children have just one measurement and 90% have at most 3 measurements. For this reason, in our analysis we assume that only a single observation per subject is available, which is chosen uniformly from the available observations. For comparison purposes, we also conducted this analysis using the full data set, with corresponding figures being relegated to the Supplementary Material [Petersen, Deoni and Müller (2019)].

We investigated the patterns of MWF development in the $p = 21$ regions/tracts listed in Table 1. A subset of the observed data are shown in the left panel of Figure 2, where the children were binned into 35 age groups of equal width, and one child per bin was chosen randomly for display. The right panel of Figure 2 also shows the preliminary estimates of the mean MWF levels, using local linear smoothing with a bandwidth of 50 days. This bandwidth was chosen by five-
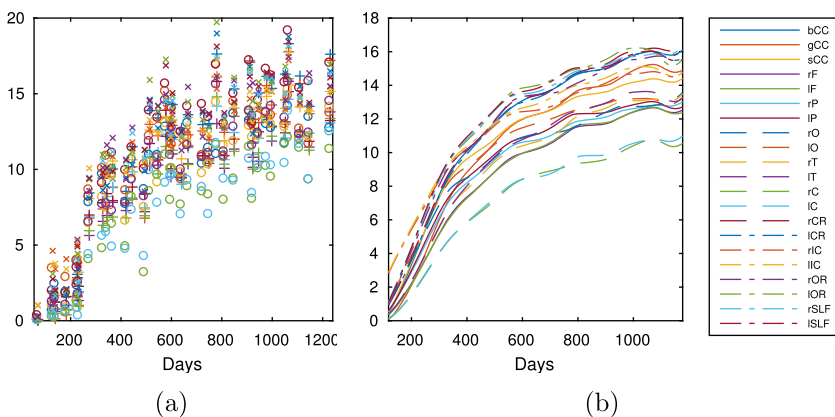

FIG. 2.   (a) *Scatterplot of MWF levels for a subset of* 20 *children, with different plotting characters and colors corresponding to distinct brain regions.* (b) *Estimated mean functions for all* 21 *regions.*

fold cross validation combined across all regions. Comparison with Figure 2 in Petersen, Deoni and Müller (2019) shows similar patterns for estimates obtained from the complete data set.

We then computed cross-sectional covariance estimates as in (2.12); the alternative estimators in (2.3) and (2.11) are not shown due to their inferior performance, which is expected theoretically and empirically demonstrated in the simulations of Section 4. For bandwidth choice, a standard cross validation criterion would not be expected to perform well as the raw covariance matrices are known to live on the boundary of the space (see Remark 2). Letting $\hat{\Sigma}_{-i}(X_i)(h)$ denote a leave-out estimate of $\Sigma(X_i)$ using bandwidth h, we consider the following criteria:

(3.1)
$$\hat{h}_1 = \underset{h}{\operatorname{argmin}} \sum_{i=1}^{n} d_F^2(\hat{C}_i, \hat{\Sigma}_{-i}(X_i)(h))^2,$$

$$\hat{h}_2 = \underset{h}{\operatorname{argmin}} \sum_{i=1}^{n} \operatorname{tr}(\hat{\Sigma}_{-i}^{+}(X_i)(h)\hat{C}_i),$$

where A^+ denotes the pseudo inverse of a matrix A. As expected, $\hat{h}_1$ tends to undersmooth the data, whereas $\hat{h}_2$ oversmooths due to the inversion. We found that using the geometric mean $\hat{h}_{\text{opt}} = (\hat{h}_1\hat{h}_2)^{1/2}$ provides a reasonable choice of 150 days for the covariance smoothing, using five-fold cross validation.

We visualize the 231 distinct covariance curves $\hat{\Sigma}_{jk}(x)$, $1 \leq j \leq k \leq 21$ for four ages in Figure 3. We see from the diagonals that variability of MWF levels increases with age, while covariability also scales but to a lesser extent. An interesting finding is the weak dependence of the right occipital region with all other regions, despite having a similar variability pattern. We can also examine the dependency patterns while normalizing the scale by considering the 210 distinct *correlation* curves $\hat{\mathcal{R}}_{jk}$, $1 \leq j < k \leq 21$, this time using Functional Principal Component Analysis (FPCA). We note here that the correlation curves are dependent as they are constructed from the same individuals so that many of the basic results such as consistency of FPCA are not applicable. We also ignore the constraint that correlation curves are between -1 and 1 when implementing FPCA. However, none of this precludes us to use FPCA for purely descriptive purposes, and this is the premise for the following FPCA application.

Figure 4(a) shows the pointwise average of these correlation curves along with a functional boxplot. The mean curve indicates that correlation between regional myelin development remains roughly constant, when averaging across all region pairs. Comparing these results with the estimates for the full data set, we again find consistency in the dynamics of covariance estimates [Figure 3 and Figure 8 in Petersen, Deoni and Müller (2019)], as well as similar patterns in the corresponding correlation curves [Figure 4 and Figure 9 in Petersen, Deoni and Müller (2019)].
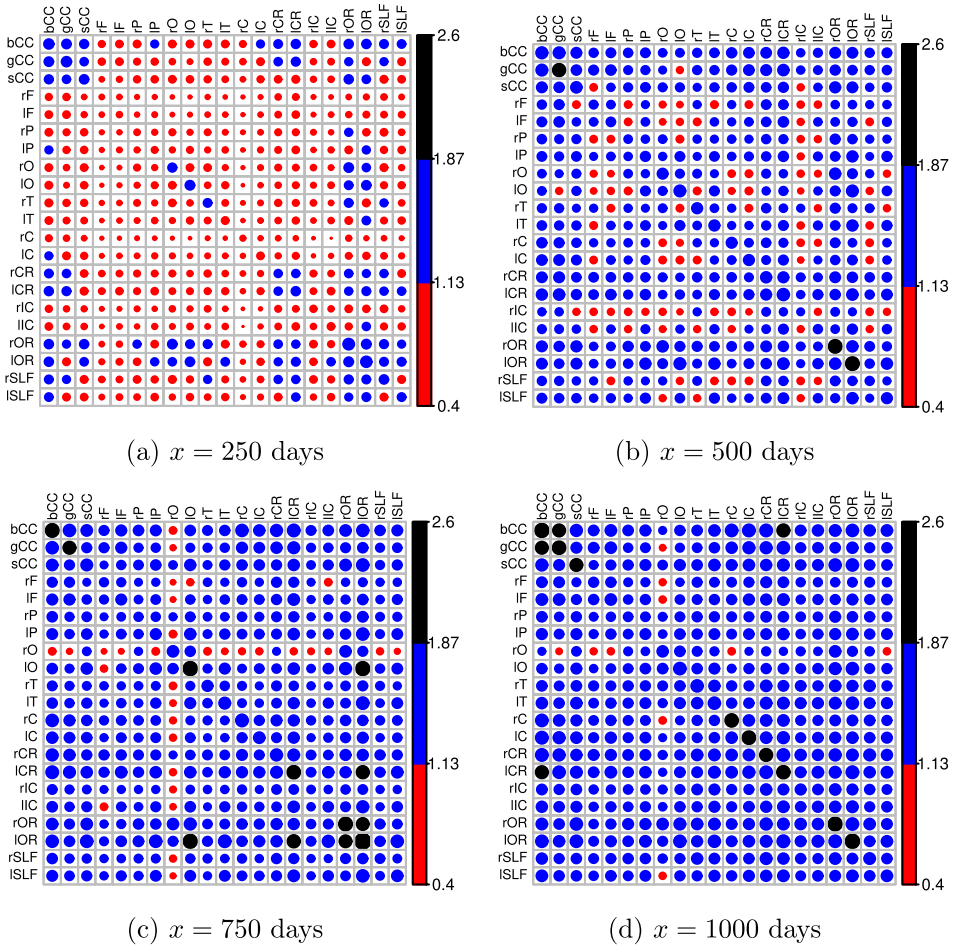
FIG. 3. *Estimated MWF covariance matrices at four distinct ages x (in days). Circle size indicates magnitude of covariance (all are positive), and the color/shade indicates a discretization of the covariance values for easier visual comparison.*

The first $k = 3$ eigenfunctions representing the principal deviations from the average correlation curve observed in the estimated correlations are shown in Figure 4(b); the fractions of variance explained (FVE) by these three curves are 90.1%, 6.15%, and 2.76%, respectively. The first eigenfunction has an increasing trend up to day 800, followed by a decreasing trend. This indicates that variability between correlation patterns amongst distinct region pairs is mostly explained by an increased separation from the mean correlation up until approximately 800 days of age, followed by a regression to the mean, that is, moving towards the average correlation. The eigenfunction curves in Figure 4(b) can be further interpreted by examining the corresponding functional principal component scores in
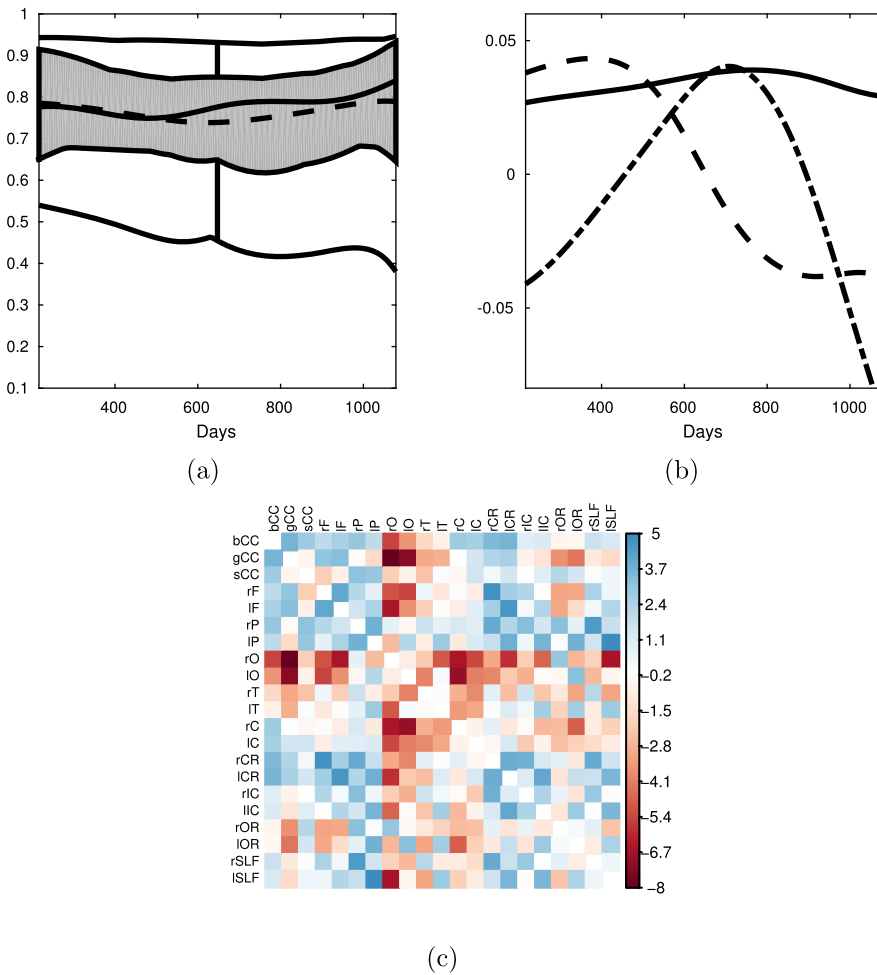
(a)

(b)

(c)

FIG. 4. (a) *Functional boxplot and mean curve* (*dashed*), *for correlation curves between the* 21 *regions*. (b) *First* (*solid*, FVE = 90.1%), *second* (*dashed*, FVE = 6.15%), *and third* (*dot-dash*, FVE = 2.76%) *eigenfunction representing the variability in the collection of estimated correlation curves*. (c) *First principal scores for all region pairs. Here FVE stands for Fraction of Variance Explained* [*for Functional Principal Component Analysis and related issues, see, e.g.,* Wang, Chiou and Müller (2016)].

Figure 4(c). As previously observed, the occipital regions have markedly lower correlations with the remaining regions as indicated by the negative scores in this component. It can also be seen that the strongest correlations in region pairs tend to include the Frontal and Parietal regions as well as the Corona Radiata.

We now aim at modeling the relation between myelin development and cognitive ability, as quantified by the ELC score, building on the above approaches. Denoting by $E_i(x)$ a latent process that tracks the ELC score of the ith child con-

tinuously over the age domain, we are able to actually observe only $E_i(X_i)$, the cognitive score at the random time X_i at which the child is studied. Recall that $U_i(x)$ is the vector latent process containing the MWF levels for the various regions across age, with mean curve $\mu(x)$. Since both cognition and myelin development are dynamic processes, we consider a varying coefficient linear model

$$(3.2) \qquad E[E_i(x)|U_i(x)] = \beta_0 + \beta^\top(x)(U_i(x) - \mu(x)),$$

where in our specific application we can substitute fixed intercept value $\beta_0 = 100$, owing to the fact that ELC scores are standardized to have a mean value of 100 and constant variance.

In view of the linear structure of (3.2), the solution for the slope vector function is

$$\beta(x) = \Sigma^{-1}(x)\Gamma(x),$$

where $\Gamma(x) = \text{Cov}(U_i(x), E_i(x))$ is the dynamic vector of covariances between myelin development and the ELC score, and $\Sigma(x)$ is the time-varying covariance matrix that we are targeting with the approach described above.

We then propose to estimate the parameter functions β in (3.2) by first estimating the elements $\Gamma_j(x)$, $j = 1, \ldots, 21$. For this, we compute

$$G_{ij} = (Y_{ij} - \hat{\mu}_j(X_i))(E_i(X_i) - 100)$$

and then apply local linear smoothing for pairs (X_i, G_{ij}) with bandwidth $h = 150$ days. Finally, we use the estimate $\hat{\Sigma}(x)$ obtained by Fréchet smoothing to compute the regularized ridge estimate

$$(3.3) \qquad \hat{\beta}_\lambda(x) = [\hat{\Sigma}(x) + \lambda I_p]^{-1} \hat{\Gamma}(x).$$

The bandwidths for the estimation of μ and Σ were set to 50 and 150, as done previously, and the ridge parameter was chosen to minimize prediction error using five-fold cross validation, resulting in $\hat{\lambda}_{\text{opt}} = 0.14$. Estimates $\hat{\Gamma}(x)$ and $\hat{\beta}_{\hat{\lambda}_{\text{opt}}}$ are shown in Figures 5(a) and 5(b). While the covariability between each MWF level and ELC score has similar patterns over time, the interdependencies between the MWF levels cause the estimates of the components of the vector function β to be considerably more complex.

For example, these estimates indicate a stark contrast in the effects of myelination level in the Genu of the Corpus Callosum, which is associated with a lower cognitive score, versus white matter in the right Parietal lobe, which is associated with a higher score, particularly pronounced at 600 days, and also between the right Optic Radiation, with lower scores associated, and left Corona Radiata, with higher scores associated and most pronounced at 800 days. Despite the aforementioned similarities in mean and covariance estimates between the reduced and full data sets, comparison of the estimates of the functional coefficients β shown in
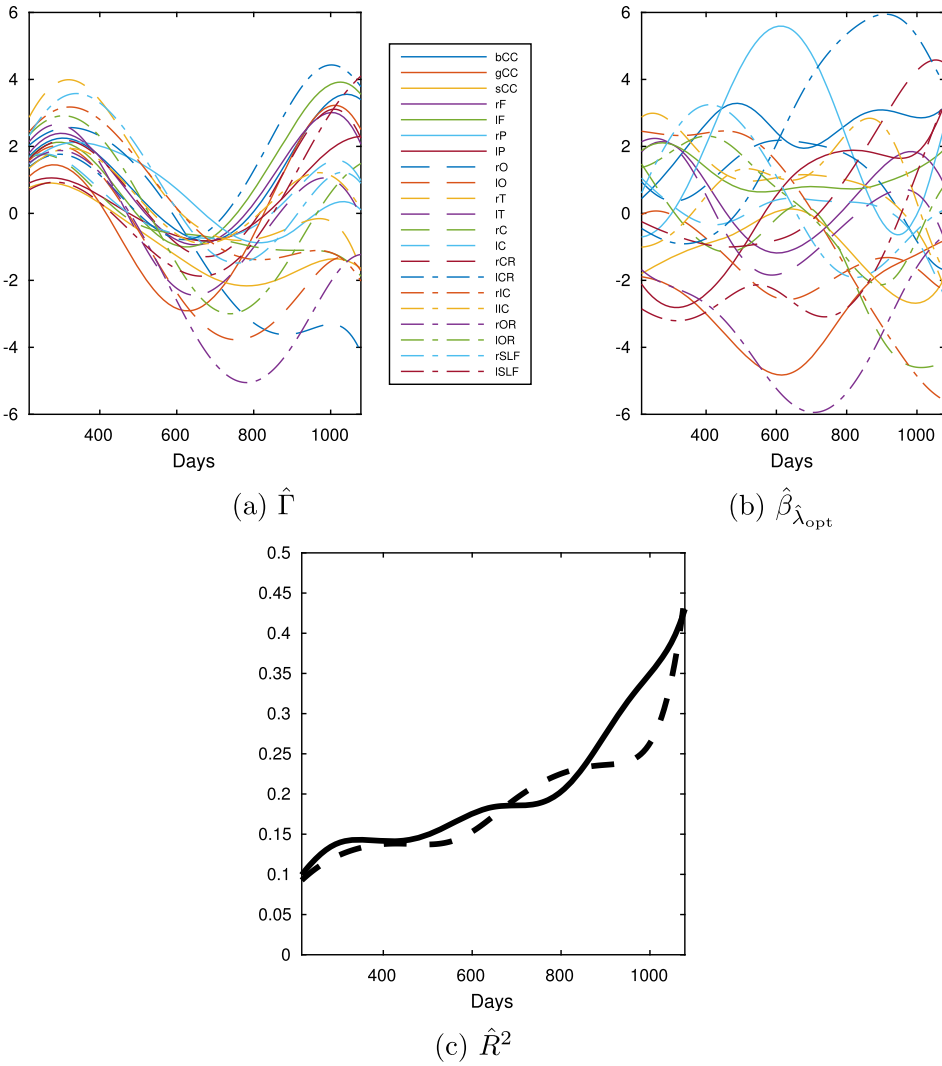
(a) $\hat{\Gamma}$

(b) $\hat{\beta}_{\hat{\lambda}_{\mathrm{opt}}}$

(c) $\hat{R}^2$

FIG. 5. *Estimates $\hat{\Gamma}(x)$ in (a), $\hat{\beta}_{\hat{\lambda}_{\mathrm{opt}}}(x)$ in (b), and $\hat{R}^2(x)$ in (c), corresponding to the varying coefficient model* (3.2) *for regressing ELC scores on myelin water fraction levels. In panel* (c), *estimates based on both one observation per subject* (*solid line*) *and the full data set* (*dashed line*) *are shown.*

Figure 5 and Figure 10 in Petersen, Deoni and Müller (2019) demonstrate their sensitivity to changes in $\hat{\Sigma}(x)$ due to the inversion in (3.3).

Another meaningful measure we can extract is an estimate of the time-varying coefficient of determination, or multiple correlation,

$$R^2(x) = \Gamma(x)^\top \Sigma(x)^{-1} \Gamma(x),$$

which we estimate by

$$\hat{R}^2(x) = \hat{\Gamma}(x)^\top [\hat{\Sigma}(x) + \hat{\lambda}_{\mathrm{opt}} I_p]^{-1} \hat{\Gamma}(x),$$

with the resulting estimates shown in Figure 5(c), where also the corresponding estimates using the complete data set are shown. One finds that while MWF levels only moderately account for variability in ELC score at younger ages, there is a sharp and steady increase in their predictive power between the ages of two and three.

These exploratory results clearly demonstrate the evolving relationships between maturing brain structure and developing cognitive functioning and behaviour, particularly throughout early neurodevelopment. In older children, adolescents, and adults, neuroimaging results generally conform to the hypothesis that increased brain structure, integrity, and myelin are associated with improved cognitive function. Our results suggest that this relationship is less clear and exhibits complex nonlinearities during infancy and early childhood. For example, our finding that reduced myelination within portions of the Corpus Callosum is associated with increased cognitive functioning runs contrary to preconceptions derived from adult studies [Fryer et al. (2008)], however, may be a natural consequence of other prior findings suggesting that slower maturation is also associated with improved cognitive outcomes [Deoni et al. (2016), Erus et al. (2014)].

With respect to our finding that brain structure and function relationships become increasingly stronger with age, this too may provide insight into the increasing importance of white matter with age, with grey matter structure playing the more important role in earlier life [Smyser et al. (2016)]. However, it is also important to note that early cognitive measures suffer high variability and poor predictive ability prior to 12–18 months of age [Slater (1997)], so that this finding may also reflect intrinsic measurement variability. Clearly, in addition to the estimation techniques introduced here, it will be useful in future work to develop inferential methods for testing and confidence sets in order to assess the strength of evidence for these findings.

4. Simulations.   The simulations reported in this section aim, first, to demonstrate the advantages of local Fréchet smoothing over the Nadaraya–Watson smoother and, second, to establish the reliability of estimates obtained from real data in Section 3.2. The simulations were conducted under the setting $T = 1$, with X_i being independently sampled from a beta distribution with shape parameters 0.5 and 1.8, which roughly approximates the shape of the age distribution in the BAMBAM data. Then, $Y_i | X_i, i = 1, \ldots, n$, were generated according to the model

(4.1) $$Y_i | X_i = \mu(X_i) + [\Sigma(X_i)]^{1/2} Z_i.$$

In order to emulate the BAMBAM data, for a fixed dimension p, the mean vector $\mu(x)$ had components $\mu_j(x) = b_j - 8(x - c_j)^2$, $j = 1, \ldots, p$, where $b_j \sim$

$\mathcal{U}(10, 16)$ and $c_j \sim \mathcal{U}(1, 1.1)$ were drawn once and fixed for all simulations. These parametric mean functions were chosen based on the estimated mean myelin trajectories (see Figure 2 in Section 3.2). The covariance $\Sigma(x)$ was formed by generating a $p \times p$ matrix A with independent $\mathcal{N}(0, 0.5)$ random variables in each entry, then computing $S = 0.5(A + A^\top)$. A second $p \times p$ matrix V was generated with elements drawn independently as $\mathcal{U}(0, 0.5)$, from which $\theta = 0.5(V + V^\top)$ was computed. Finally, with Exp denoting matrix exponentiation and $\odot$ the Hadamard product, we formed

$$(4.2) \qquad \Sigma(x) = (1 + 10x + 20x^5) \operatorname{Exp}\big[S \odot \sin\big(2\pi\theta(x + 0.1)\big)\big].$$

The polynomial factor emulates the increasing variability of MWF levels seen in the BAMBAM data, with minor fluctuations induced by the sine component. Repeated simulations were run by sampling the random variate Z_i independently of X_i as a standard p-dimensional Gaussian random vector. Mean trajectory estimates $\hat{\mu}_j$ were obtained using local linear regression with the bandwidth chosen by ordinary leave-one-out cross validation, pooled across j, and separately for each simulation run. Covariance estimation was performed using both Nadaraya–Watson (2.11), local Fréchet (2.12), and standard local linear estimators, which are obtained as in (2.3), with $m = 1$ and weights $w_{in,1}$ as specified in (2.5). For the latter, the various bandwidths were all taken to be equal, that is, $h_{jk} \equiv h$.

To compare the various estimates, for each fixed bandwidth h over a grid, we computed the integrated squared error (ISE)

$$\text{ISE}(h) = \int_0^1 d_F^2\big(\hat{\Sigma}(x)(h), \Sigma(x)\big) \, dx$$

for each simulation run, where $\hat{\Sigma}(x)(h)$ stands for a generic estimate of $\Sigma(x)$ using bandwidth h. The results for comparing (2.11), (2.12), and (2.3) using this metric are in Table 2 for 100 simulations, dimensions $p = 20, 40$, and sample

TABLE 2
Logarithms of average ISE for Nadaraya–Watson (2.11), *ordinary local linear* (2.3), *and local Fréchet* (2.12) *estimators, minimized over a grid of bandwidth values h. The minimizing bandwidth is given in parentheses*

Dimension	Method	$n = 250$	$n = 500$	$n = 1000$
$p = 20$	Nadaraya–Watson	11.981 (0.08)	11.958 (0.06)	11.937 (0.05)
	Local linear	10.799 (0.43)	10.430 (0.33)	10.032 (0.29)
	Local Fréchet	10.798 (0.43)	10.430 (0.33)	10.031 (0.29)
$p = 40$	Nadaraya–Watson	14.693 (0.09)	14.673 (0.07)	14.655 (0.05)
	Local linear	13.846 (0.60)	13.457 (0.33)	13.068 (0.29)
	Local Fréchet	13.845 (0.60)	13.456 (0.37)	13.068 (0.29)

sizes $n = 250, 500, 1000$. The settings we investigated include parameter values that approximate the BAMBAM data, and go beyond by increasing both dimension and sample size. One finds that average ISE values are uniformly lower for the local linear and Fréchet methods, and the very small optimal bandwidths used for the Nadaraya–Watson estimator particularly reflect its well-known inherent bias issues, and increasing sample size leads only to very small declines or no declines in ISE, in contrast to the local linear and Fréchet methods, where increasing sample sizes lead to noticeable improvements. While the tabulated *integrated* MSE results demonstrate only slight improvement using local Fréchet estimation versus ordinary local linear estimation, in fact the *pointwise* discrepancies $d_F(\hat{\Sigma}(x)(h), \Sigma(x))$ are uniformly lower for the local Fréchet method, with larger discrepancies near the boundaries where the ordinary local linear estimate tends to suffer from influential negative eigenvalues.

The simulation results are visualized in Figure 6, demonstrating the ISE values (in log scale) over all simulation runs rather than just the mean, where we leave out the ordinary local linear estimator. The local Fréchet approach clearly outperforms the Nadaraya–Watson estimator uniformly over all settings, which, again due to its bias, is seen not to display the improved performance with increasing sample size n that is found for the local Fréchet estimator and is expected for a reasonable estimation approach.

To further validate our analysis in Section 3, we implemented two additional estimates of the dynamic covariance for the setting $n = 250$ and $p = 20$. The first alternative estimate is the local Fréchet estimator under the square root metric
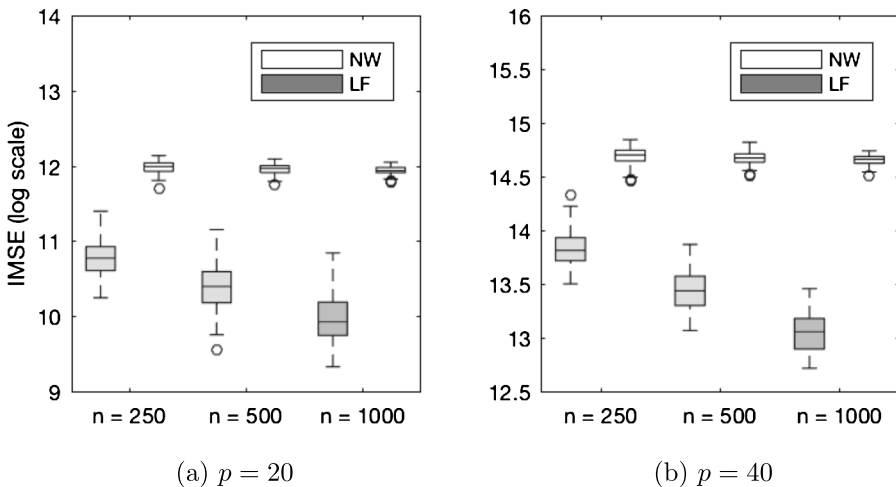

(a) $p = 20$ (b) $p = 40$

FIG. 6. *Boxplots of integrated squared errors for Nadaraya–Watson (NW) and local Fréchet (LF) estimators for $p = 20, 40$ and $n = 250, 500, 1000$, using the optimal bandwidths given in Table 2.*

$$d(C_1, C_2) = d_F(C_1^{1/2}, C_2^{1/2}),$$

$$\hat{\Sigma}_{\oplus}^d(x) = \left(\sum_{i=1}^{n} w_{in,1}(x, h) \hat{C}_i^{1/2} \right)^2.$$

The appeal of this estimate is that it is an intrinsic estimator that does not require projection onto $\mathcal{S}_p$. However, as noted in Section 2.2, this estimator no longer targets the ordinary covariance, but rather the d-covariance of Tavakoli et al. (2016) in (2.10). This inherent bias resulted in a worse average ISE of 11.376 (on the log scale, and minimized over a grid of bandwidths) when compared to local Fréchet estimation under the Frobenius metric in Table 2, although the performance was better than that of the standard Nadaraya–Watson estimate.

A second alternative estimate was considered in order to assess the relative efficiency of using just one observation per subject versus the full data set in the BAMBAM analysis. This estimate was obtained by generating repeated observations for the $n = 250$ subjects and then computing the local Fréchet estimate using all observations, ignoring dependencies. To mimic the BAMBAM data, we generated independent obervation counts $N_i \in \{1, 2, 3, 4\}$ with probabilities $(0.48, 0.28, 0.14, 0.1)$, resulting in a total of 497 observations including repeats, and these numbers were used across all simulation runs. For indices i with $N_i > 1$, this required an adaptation of the generative model in (4.1) in order to include dependencies. This was done by generating independent timepoints X_{ij}, $j = 1, \ldots, N_i$, according to the beta distribution as before, followed by a zero-mean multivariate normal random vector $(Z_{i1}^\top, \ldots, Z_{iN_i}^\top)^\top$ with $\text{Var}(Z_{ij}) = I_p$ and $\text{Cov}(Z_{ij}, Z_{ij'}) = 0.2 I_p$, $j, j' = 1, \ldots, N_i$, $j' \neq j$. Lastly, the random vector

$$Y_{ij}|X_{ij} = \mu(X_{ij}) + [\Sigma(X_{ij})]^{1/2} Z_{ij}$$

was computed. Unsurprisingly, the increased information resulted in an improved average ISE of 10.465 (again in log scale). Hence, when repeats are available, they can indeed be beneficial. However, the methodology does not require repeats and, most importantly, the challenges mentioned in Section 2.2 associated with single observation data are not alleviated for very sparse longitudinal data such as those available in the BAMBAM study.

5. Theoretical justifications. We analyze the pointwise behavior of the estimate $\hat{\Sigma}^{\text{LF}}(x)$ and establish the rate of convergence for the metric $d_F(\hat{\Sigma}^{\text{LF}}(x), \Sigma(x))$. If μ were known, so that C_i in (2.9) can be computed, the estimator

$$(5.1) \qquad \tilde{\Sigma}^{\text{LF}}(x) = \underset{C \in \mathcal{S}_p}{\arg\min} \sum_{i=1}^{n} w_{in,1}(x, h) d_F^2(C, C_i)$$

corresponds to the estimator studied in Petersen and Müller (2018). Thus, we can make use of the inequality

$$d_F(\hat{\Sigma}^{\text{LF}}(x), \Sigma(x)) \leq d_F(\hat{\Sigma}^{\text{LF}}(x), \tilde{\Sigma}^{\text{LF}}(x)) + d_F(\tilde{\Sigma}^{\text{LF}}(x), \Sigma(x)).$$

Suppose that the marginal densities f_X of X_1 and $f_U(\cdot, x)$ of $U_1(x)$ exist, the latter for all $x \in [0, T]$. Let $\|\cdot\|$ be the standard Euclidean norm on $\mathbb{R}^p$, and $h = h_n \to 0$ be a bandwidth sequence. We require the following assumptions.

(A1) The kernel function K is a symmetric probability density function with support $[-1, 1]$.

(A2) The density f_X of X_1 is twice continuously differentiable and $\inf_{x \in [0, T]} f(x) > 0$.

(A3) There is a constant $M_1 > 0$ such that $\sup_{x \in [0, T]} \|U_1(x)\| \le M_1$ almost surely. Futhermore,

$$\inf_{x \in [0, T], \|y\| \le M_1} f_U(y, x) > 0.$$

(A4) The function $f_U(u, x)$ is continuous with respect to both arguments and twice continuously differentiable with respect to x on the interior.

(A5) For any $x \in [0, T]$ and a null sequence $q_n = o(1)$, the auxiliary estimates $\hat{\mu}_j$ satisfy

$$\sup_{|y - x| \le h} |\hat{\mu}_j(y) - \mu_j(y)| = O_p(q_n).$$

Assumption (A1) is common for estimation of a function with bounded support, but can be relaxed by controlling the tail behavior of K. Assumption (A2) is also widely used in local polynomial estimation settings, while (A3) is a natural assumption which has the implication that the conditional density of $Y_1|X_1$ is well behaved. Assumption (A4) can also be relaxed by controlling the pointwise tail behavior of the process U_1. The required rate in (A5) is needed to control the error incurred by using $\hat{C}_i$ in place of C_i in the estimation, where the upper bound $q = n^{-2/5} \log(n)$ is known to hold for local linear estimators under certain assumptions, which include that the functions $\mu_j(x)$ are twice continuously differentiable and conditions on the bandwidth [Mack and Silverman (1982), Fan and Gijbels (1996)].

THEOREM 1. *Let $h = h_n \to 0$ and $nh \to \infty$. Under assumptions* (A1)–(A5), *for any interior point $x \in (0, T)$, the local Fréchet estimator satisfies*

$$d_F(\Sigma(x), \hat{\Sigma}^{\mathrm{LF}}(x)) = O_p(h^2 + (nh)^{-1/2} + q_n).$$

COROLLARY 1. *Let $\hat{\mathcal{R}}^{\mathrm{LF}}(x)$ and $\mathcal{R}(x)$ be the correlation matrices corresponding to $\hat{\Sigma}^{\mathrm{LF}}(x)$ and $\Sigma(x)$, respectively. Under the assumptions of Theorem* 1, *for any $x \in (0, T)$,*

$$d_F(\mathcal{R}(x), \hat{\mathcal{R}}^{\mathrm{LF}}(t)) = O_p(h^2 + (nh)^{-1/2} + q_n).$$

From the theorem and corollary, the bandwidth choice which leads to the fastest rate of convergence is $h = h(n) \sim n^{-1/5}$ and the resulting O_p rate is $n^{-2/5} + q_n$. The first term corresponds to the optimal rate of pointwise convergence for nonparametric estimation of regression functions that are twice continuously differentiable. A proof of Theorem 1 is provided below, while that of the corollary is straightforward and is omitted.

PROOF OF THEOREM 1.   First, assumptions (A1)–(A4) can be used to verify the conditions necessary to invoke Theorems 3 and 4 of Petersen and Müller (2018), with the exception that the space $\mathcal{S}_p$ is not bounded. With minor alterations to the proofs, assumption (A3) implies that these continue to hold, since $\mathcal{S}_p$ is a subset of a Euclidean space, implying

$$d_F\big(\Sigma(x), \tilde{\Sigma}^{\mathrm{LF}}(x)\big) = O_p\big(h^2 + (nh)^{-1/2}\big).$$

By the contraction property of the projection $\Pi_{\mathcal{S}_p}$, it follows that

$$d_F\big(\tilde{\Sigma}^{\mathrm{LF}}(x), \hat{\Sigma}^{\mathrm{LF}}(x)\big) \leq \frac{1}{n}\left\|\sum_{i=1}^{n} w_{in,1}(x)(C_i - \hat{C}_i)\right\|_F,$$

where $\|\cdot\|_F$ is the Frobenius norm. Assumption (A5) also implies that, for large enough n, we have

$$\|C_i - \hat{C}_i\|_F \leq 4M_1\|\hat{\mu}(X_i) - \mu(X_i)\|.$$

Under (A1), it can easily be shown that $|w_{in,1}(x)| = O_p(K_h(X_i - x))$, where the O_p term is uniform over i. Hence, using (A5), we have

$$d_F\big(\tilde{\Sigma}^{\mathrm{LF}}(x), \hat{\Sigma}^{\mathrm{LF}}(x)\big) \leq \frac{1}{n}\sum_{i=1}^{n}|w_{in,1}(x)|\|C_i - \hat{C}_i\|_F$$

$$= O_p\left(\frac{q_n}{n}\sum_{i=1}^{n}K_h(X_i - x)\right) = O_p(q_n). \qquad \square$$

SUPPLEMENTARY MATERIAL

Supplement to "Fréchet estimation of time-varying covariance matrices from sparse data, with application to the regional co-evolution of myelination in the developing brain" (DOI: 10.1214/18-AOAS1195SUPP; .pdf). The Supplementary Material contains figures similar to Figures 2–5 in the paper, but corresponding to results obtained by using the full data set of 440 observations, including repeated measurements for approximately half of the 223 total subjects.

REFERENCES

ŞENTÜRK, D. and MÜLLER, H.-G. (2010). Functional varying coefficient models for longitudinal data. *J. Amer. Statist*. *Assoc*. **105** 1256–1264. MR2752619

ŞENTÜRK, D. and NGUYEN, D. V. (2011). Varying coefficient models for sparse noise-contaminated longitudinal data. *Statist*. *Sinica* **21** 1831–1856. MR2896001

ASH, R. B. and GARDNER, M. F. (1975). *Topics in Stochastic Processes*. *Probability and Mathematical Statistics* **27**. Academic Press, New York. MR0448463

BARTZOKIS, G., LU, P., TINGUS, K., MENDEZ, M., RICHARD, A., PETERS, D. G. et al. (2010). Lifespan trajectory of myelin integrity and maximum motor speed. *Neurobiol*. *Aging* **31** 1554–1562.

BILENBERG, N. (1999). The child behavior checklist (CBCL) and related material: Standardization and validation in Danish population based and clinically based samples. *Acta Psychiatr*. *Scand*. **100** 2–52.

BRODY, B. A., KINNEY, H. C., KLOMAN, A. S. and GILLES, F. H. (1987). Sequence of central nervous system myelination in human infancy. I. An autopsy study of myelination. *J. Neuropathol*. *Exp*. *Neurol*. **46** 283–301.

CARMICHAEL, O., CHEN, J., PAUL, D. and PENG, J. (2013). Diffusion tensor smoothing through weighted Karcher means. *Electron*. *J. Stat*. **7** 1913–1956. MR3084676

CASEY, B., GALVAN, A. and HARE, T. A. (2005). Changes in cerebral functional organization during cognitive development. *Curr*. *Opin*. *Neurobiol*. **15** 239–244.

CHIOU, J.-M. and MÜLLER, H.-G. (2016). A pairwise interaction model for multivariate functional and longitudinal data. *Biometrika* **103** 377–396. MR3509893

CHLEBOWSKI, C., ROBINS, D. L., BARTON, M. L. and FEIN, D. (2013). Large-scale use of the modified checklist for autism in low-risk toddlers. *Pediatrics* **131** e1121–e1127.

DAI, X., MÜLLER, H.-G., WANG, J.-L. and DEONI, S. C. (2017a). Age-dynamic networks and functional correlation for early white matter myelination. Preprint.

DAI, X., HADJIPANTELIS, P., WANG, J.-L., DEONI, S. C. and MÜLLER, H.-G. (2017b). Longitudinal associations between white matter maturation and cognitive development across early childhood. Preprint.

DEONI, S. C. (2007). High-resolution T1 mapping of the brain at 3T with driven equilibrium single pulse observation of T1 with high-speed incorporation of RF field inhomogeneities (DESPOT1-HIFI). *J. Magn*. *Reson*. *Imaging* **26** 1106–1111.

DEONI, S. C. (2011). Correction of main and transmit magnetic field (B0 and B1) inhomogeneity effects in multicomponent-driven equilibrium single-pulse observation of T1 and T2. *Magn*. *Reson*. *Med*. **65** 1021–1035.

DEONI, S. C. and KOLIND, S. H. (2015). Investigating the stability of mcDESPOT myelin water fraction values derived using a stochastic region contraction approach. *Magn*. *Reson*. *Med*. **73** 161–169.

DEONI, S. C., PETERS, T. M. and RUTT, B. K. (2004). Determination of optimal angles for variable nutation proton magnetic spin-lattice, T1, and spin-spin, T2, relaxation times measurement. *Magn*. *Reson*. *Med*. **51** 194–199.

DEONI, S. C., RUTT, B. K. and PETERS, T. M. (2003). Rapid combined T1 and T2 mapping using gradient recalled acquisition in the steady state. *Magn*. *Reson*. *Med*. **49** 515–526.

DEONI, S. C., RUTT, B. K. and PETERS, T. M. (2006). Synthetic T1-weighted brain image generation with incorporated coil intensity correction using DESPOT1. *Magn*. *Reson*. *Imaging* **24** 1241–1248.

DEONI, S. C., WARD, H. A., PETERS, T. M. and RUTT, B. K. (2004). Rapid T2 estimation with phase-cycled variable nutation steady-state free precession. *Magn*. *Reson*. *Med*. **52** 435–439.

DEONI, S. C., WILLIAMS, S. C., JEZZARD, P., SUCKLING, J., MURPHY, D. G. and JONES, D. K. (2008a). Standardized structural magnetic resonance imaging in multicentre studies using quantitative T1 and T2 imaging at 1.5 T. *NeuroImage* **40** 662–671.

DEONI, S. C., RUTT, B. K., ARUN, T., PIERPAOLI, C. and JONES, D. K. (2008b). Gleaning multi-component T1 and T2 information from steady-state imaging data. *Magn. Reson. Med.* **60** 1372–1387.

DEONI, S. C., MERCURE, E., BLASI, A., GASSTON, D., THOMSON, A., JOHNSON, M., WILLIAMS, S. C. and MURPHY, D. G. (2011). Mapping infant brain myelination with magnetic resonance imaging. *J. Neurosci.* **31** 784–791.

DEONI, S. C., O'MUIRCHEARTAIGH, J., ELISON, J. T., WALKER, L., DOERNBERG, E., WASKIEWICZ, N., DIRKS, H., PIRYATINSKY, I., DEAN III, D. C. and JUMBE, N. (2016). White matter maturation profiles through early childhood predict general cognitive ability. *Brain Struct. Funct.* **221** 1189–1203.

DEONI, S. C., DEAN, D. C., REMER, J., DIRKS, H. and O'MUIRCHEARTAIGH, J. (2015). Cortical maturation and myelination in healthy toddlers and young children. *NeuroImage* **115** 147–161.

ERUS, G., BATTAPADY, H., SATTERTHWAITE, T. D., HAKONARSON, H., GUR, R. E., DAVATZIKOS, C. and GUR, R. C. (2014). Imaging patterns of brain development and their relationship to cognition. *Cereb. Cortex* **25** 1676–1684.

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. Chapman & Hall, London. MR1383587

FIELDS, R. (2005). Myelination: An overlooked mechanism of synaptic plasticity? *Neuroscientist* **11** 528–531.

FIELDS, R. D. (2008). White matter in learning, cognition and psychiatric disorders. *Trends Neurosci.* **31** 361–370.

FLYNN, S., LANG, D., MACKAY, A., GOGHARI, V., VAVASOUR, I., WHITTALL, K., SMITH, G., ARANGO, V., MANN, J., DWORK, A. et al. (2003). Abnormalities of myelination in schizophrenia detected in vivo with MRI, and post-mortem with analysis of oligodendrocyte proteins. *Mol. Psychiatry* **8** 811–820.

FORNARI, E., KNYAZEVA, M. G., MEULI, R. and MAEDER, P. (2007). Myelination shapes functional activity in the developing brain. *NeuroImage* **38** 511–518.

FRYER, S. L., FRANK, L. R., SPADONI, A. D., THEILMANN, R. J., NAGEL, B. J., SCHWEINSBURG, A. D. and TAPERT, S. F. (2008). Microstructural integrity of the corpus callosum linked with neuropsychological performance in adolescents. *Brain Cogn.* **67** 225–233.

GRYDELAND, H., WALHOVD, K. B., TAMNES, C. K., WESTLYE, L. T. and FJELL, A. M. (2013). Intracortical myelin links with performance variability across the human lifespan: Results from T1- and T2-weighted MRI myelin mapping and diffusion tensor imaging. *J. Neurosci.* **33** 18618–18630.

HAGMANN, P., SPORNS, O., MADAN, N., CAMMOUN, L., PIENAAR, R., WEDEEN, V. J., MEULI, R., THIRAN, J.-P. and GRANT, P. (2010). White matter maturation reshapes structural connectivity in the late developing human brain. *Proc. Natl. Acad. Sci. USA* **107** 19067–19072.

HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis*, *with an Introduction to Linear Operators*. Wiley, Chichester. MR3379106

ISHIBASHI, T., DAKIN, K., STEVENS, B., LEE, P., KOZLOV, S. and STEWART, C. E. A. (2006). Astrocytes promote myelination in response to electrical impulses. *Neuron* **49** 823–832.

KNICKMEYER, R. C., GOUTTARD, S., KANG, C., EVANS, D., WILBER, K., SMITH, J. K., HAMER, R. M., LIN, W., GERIG, G. and GILMORE, J. H. (2008). A structural MRI study of human brain development from birth to 2 years. *J. Neurosci.* **28** 12176–12182.

LANG, D., YIP, E., MACKAY, A., THORNTON, A., VILA-RODRIGUEZ, F. and MACEWAN, G. E. A. (2003). Abnormalities of myelination in schizophrenia detected in vivo with MRI, and post-mortem with analysis of oligodendrocyte proteins. *Mol. Psychiatry* **8** 811–820.

LEBEL, C. and BEAULIEU, C. (2011). Longitudinal development of human brain wiring continues from childhood into adulthood. *J. Neurosci.* **31** 10937–10947.

LENROOT, R. K. and GIEDD, J. N. (2006). Brain development in children and adolescents: Insights from anatomical magnetic resonance imaging. *Neurosci. Biobehav. Rev.* **30** 718–729.

MACK, Y. P. and SILVERMAN, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. Verw. Gebiete* **61** 405–415. MR0679685

MILLER, D. J., DUKA, T., STIMPSON, C. D., SCHAPIRO, S. J., BAZE, W. B., MCARTHUR, M. J., FOBBS, A. J., SOUSA, A. M., ŠESTAN, N., WILDMAN, D. E. et al. (2012). Prolonged myelination in human neocortical evolution. *Proc. Natl. Acad. Sci. USA* **109** 16480–16485.

MULLEN, E. M. et al. (1995). *Mullen Scales of Early Learning*. AGS, Circle Pines, MN.

NAGY, Z., WESTERBERG, H. and KLINGBERG, T. (2004). Maturation of white matter is associated with the development of cognitive functions during childhood. *J. Cogn. Neurosci.* **16** 1227–1233.

PAUS, T., COLLINS, D., EVANS, A., LEONARD, G., PIKE, B. and ZIJDENBOS, A. (2001). Maturation of white matter in the human brain: A review of magnetic resonance studies. *Brain Res. Bull.* **54** 255–266.

PETERSEN, A., DEONI, S. and MÜLLER, H.-G. (2019). Supplement to "Fréchet estimation of time-varying covariance matrices from sparse data, with application to the regional co-evolution of myelination in the developing brain." DOI:10.1214/18-AOAS1195SUPP.

PETERSEN, A. and MÜLLER, H.-G. (2016). Fréchet integration and adaptive metric selection for interpretable covariances of multivariate functional data. *Biometrika* **103** 103–120. MR3465824

PETERSEN, A. and MÜLLER, H.-G. (2018). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* To appear. Available at arXiv:1608.03012.

PIGOLI, D., ASTON, J. A. D., DRYDEN, I. L. and SECCHI, P. (2014). Distances and inference for covariance operators. *Biometrika* **101** 409–422. MR3215356

RODIER, P. M. (1995). Developing brain as a target of toxicity. *Environ. Health Perspect.* **103** 73–76.

SHAFEE, R., BUCKNER, R. L. and FISCHL, B. (2015). Gray matter myelination of 1555 human brains using partial volume corrected MRI images. *NeuroImage* **105** 473–485.

SHAW, P., GREENSTEIN, D., LERCH, J., CLASEN, L., LENROOT, R., GOGTAY, N., EVANS, A., RAPOPORT, J. and GIEDD, J. (2006). Intellectual ability and cortical development in children and adolescents. *Nature* **440** 676–679.

SHAW, P., KABANI, N. J., LERCH, J. P., ECKSTRAND, K., LENROOT, R., GOGTAY, N., GREENSTEIN, D., CLASEN, L., EVANS, A., RAPOPORT, J. L. et al. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *J. Neurosci.* **28** 3586–3594.

SLATER, A. (1997). Can measures of infant habituation predict later intellectual ability? *Arch. Dis. Child.* **77** 474–476.

SMYSER, T. A., SMYSER, C. D., ROGERS, C. E., GILLESPIE, S. K., INDER, T. E. and NEIL, J. J. (2016). Cortical gray and adjacent white matter demonstrate synchronous maturation in very preterm infants. *Cereb. Cortex* **26** 3370–3378.

STILES, J. and JERNIGAN, T. (2010). The basics of brain development. *Neuropsychol. Rev.* **20** 327–348.

TAVAKOLI, S., PIGOLI, D., ASTON, J. A. and COLEMAN, J. (2016). Spatial modeling of object data: Analysing dialect sound variations across the UK. Preprint. Available at arXiv:1610.10040.

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

WOLFF, J. J., GU, H., GERIG, G., ELISON, J. T., STYNER, M., GOUTTARD, S., BOTTERON, K. N., DAGER, S. R., DAWSON, G., ESTES, A. M. et al. (2012). Differences in white matter fiber tract development present from 6 to 24 months in infants with autism. *Am. J. Psychiatr.* **169** 589–600.

XIAO, Z., QIU, T., KE, X., XIAO, X., XIAO, T., LIANG, F., ZOU, B., HUANG, H., FANG, H., CHU, K. et al. (2014). Autism spectrum disorder as early neurodevelopmental disorder: Evidence from the brain imaging abnormalities in 2–3 years old toddlers. *J. Autism Dev. Disord.* **44** 1633–1640.

YUAN, Y., ZHU, H., LIN, W. and MARRON, J. S. (2012). Local polynomial regression for symmetric positive definite matrices. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 697–719. MR2965956

ZHOU, L., HUANG, J. Z. and CARROLL, R. J. (2008). Joint modelling of paired sparse functional data using principal components. *Biometrika* **95** 601–619. MR2443178

A. PETERSEN
DEPARTMENT OF STATISTICS
  AND APPLIED PROBABILITY
UNIVERSITY OF CALIFORNIA, SANTA BARBARA
SANTA BARBARA, CALIFORNIA 93106-3110
USA
E-MAIL: petersen@pstat.ucsb.edu

S. DEONI
DEPARTMENT OF PEDIATRICS
BROWN UNIVERSITY
PROVIDENCE, RHODE ISLAND 02912
USA
E-MAIL: sdeoni@brown.edu

H.-G. MÜLLER
DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, DAVIS
DAVIS, CALIFORNIA 95616
USA
E-MAIL: hgmueller@ucdavis.edu