Modeling Functional Data With Spatially Heterogeneous Shape Characteristics

Ana-Maria Staicu* Ciprian M. Crainiceanu† Daniel S. Reich‡ David Ruppert§

October 11, 2010

Abstract

We propose a novel class of models for functional data exhibiting skewness or other shape characteristics that vary with spatial or temporal location. We use copulas so that the marginal distributions and the dependence structure can be modeled independently. Dependence is modeled with a Gaussian or t-copula, so that there is an underlying latent Gaussian process. We model the marginal distributions using the skew t family. The mean, variance, and shape parameters are modeled nonparametrically as functions of location. A computationally tractable inferential framework for estimating heterogeneous asymmetric or heavy-tailed marginal distributions is introduced. This framework provides a new set of tools for increasingly complex data collected in medical and public health studies. Our methods were motivated by and are illustrated with a state-of-the-art study of neuronal tracts in multiple sclerosis patients and healthy controls. Using the tools we have developed, we were able to find those locations along the tract most affected by the disease. However, our methods are general and highly relevant to many functional data sets. In addition to the application to one-dimensional tract profiles illustrated here, higher-dimensional extensions of the methodology could have direct applications to other biological data including functional and structural MRI.

*Assistant Professor, Department of Statistics, North Carolina State University, 2311 Stinson Drive, Campus Box 8203, Raleigh, NC 27695-8203 USA (email: ana-maria_staicu@ncsu.edu)

†Associate Professor, Department of Biostatistics, Johns Hopkins University, 615 N. Wolfe Street, E3636 Baltimore, MD, 21205, USA (email: ccrainic@jhsph.edu). Research supported by NIH grant R01NS060910.

‡Professor, Department of Neurology, Johns Hopkins University, 615 N. Wolfe Street, E3636 Baltimore, MD, 21205, USA (email: dsreich@gmail.com). Supported by the Intramural Research Program of the National Institute of Neurological Disorders and Stroke.

§Andrew Schultz, Jr., Professor of Engineering, School of Operational Research and Information Engineering, Comstock Hall, Cornell University, New York 14853-3801 USA (email: dr24@cornell.edu). Research supported by NIH grant R01NS060910.

Some Key Words: Gaussian and t-copulas; Quantile modeling; Skewed functional data; Tractography data;

1 Introduction

Most models used in functional data analysis (FDA) focus on modeling the mean and covariance functions. Although such models are often adequate, there are important cases where interest is focused on skewness or other shape characteristics. For example, the right panel of Figure 1 illustrates profiles of multiple sclerosis (MS) patients, recorded along the corpus callosum tract of the brain. The functions have skewed pointwise distributions, with the amount of skewness varying spatially. Modeling data with such characteristics represents the main focus and novelty of our work.

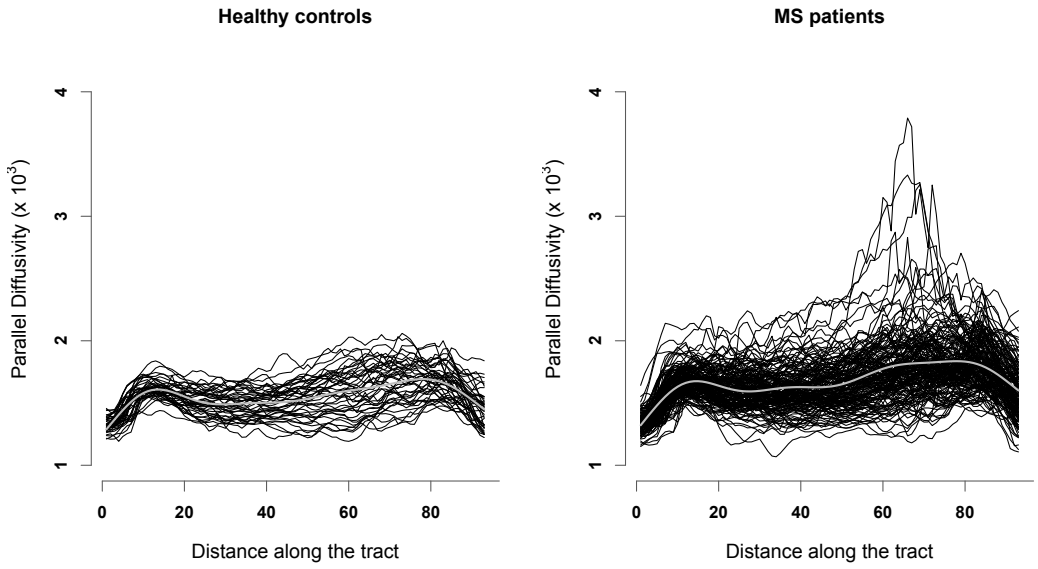

Figure 1: Parallel diffusivity profiles within the corpus callosum tract in 162 MS patients (right panel) and 42 healthy controls subjects (left panel). Penalized spline estimates of the mean functions are shown as thick grey curves. Distance units are arbitrary.

In this paper, we develop models and inferential methods for the analysis of functional skewed data. Fundamentally different from the common method of moment-based approaches of modeling functional data (Di et al., 2009; Crainiceanu et al, 2010; Staicu, Crainiceanu, and Carroll, 2010 etc.), our methods are based on pointwise distributional assumptions of the underlying stochastic process. The parametric assumptions are used to capture higher order moments of the pointwise distributions; by comparison the non-parametric approaches use only the first two moments. Specifically, our approach assumes that the pointwise distributions of the underlying stochastic process are in a parametric family with one or more shape parameters, for example, a skew normal or skew t family. The shape parameters, as well as the mean and variance, are modeled non-parametrically as function of location.

The third novel feature here is the use of copulas to model spatial dependence. We use Gaussian or t-copulas, since, unlike Archimedean copulas, they include correlation functions which are fundamental to FDA. The use of copulas allows a modular approach where the marginal distributions and dependencies can be modeled separately. In our model, the observed functions are transformations of an underlying latent Gaussian processes and, for a t-copula, a chi-squared random variable that induces tail dependence. Principal components analysis (PCA) and a Karhunen-Loève expansion can be applied to the latent Gaussian process, much as they would be applied if the observed data were Gaussian.

1.1 The diffusion tensor imaging study

The methodology described in this paper is motivated by a state-of-the-art diffusion tensor imaging (DTI) study that analyzes and compares white matter tracts in healthy individuals and MS patients. DTI is an in vivo magnetic resonance technique for imaging the white matter tracts in the brain by measuring the three-dimensional directions of water diffusion (Basser, et al. 1994, 2000). In the DTI study, we are interested in modeling DTI profiles, such as mean diffusivity and parallel diffusivity, along the corpus callosum, the major tract connecting the two cerebral hemispheres. Mean diffusivity is an orientation-independent measure of water movement within the tract, whereas parallel diffusivity estimates the component of that movement co-axial with the tract's axon bundles.

Figure 1 displays parallel diffusivity profiles for healthy controls and MS patients. Parallel

diffusivity is recorded at many locations along the tract, so that the tract profile can be viewed as a continuous curve sampled at a discrete set of points. Scientific interest includes comparing the means, quantiles, and correlation functions of the two samples; see Greven, et al. (2010) for more details on the scientific questions. At a quick look, the control and MS data seem to differ in few ways, such as mean and variability along the tract, but perhaps even more so with respect to skewness. The MS subjects' profiles exhibit skewness compared to the controls, especially around locations 50–80 along the corpus callosum tract.

These data suggest that there is a need for methods that can accommodate strong temporally- or spatially-varying skewness and within-function correlation. The computational speed is essential, since FDA is faced with continuously larger and more complex data sets. Moreover, one needs inferential methods, for example, to test whether skewness differences between spatial locations, or between groups of subjects, are statistically significant. In addition to the application to one dimensional tract profiles illustrated here, higher-dimensional extensions of the methodology could have direct applications to other biological data including functional and structural MRI. In these applications, accurate modeling of the tails of distributions is a critical step in identifying abnormalities as well as limiting false positive results.

1.2 Current methods

A reasonable approach for the analysis of tractography data is functional data analysis (FDA)—see, for example, the excellent monographs of Ramsay and Silverman (2005), Ferraty and Vieu (2006), and Ramsay, Hooker, and Graves (2009) and a journal literature too extensive to list here. A fundamental idea in FDA is to decompose the space of curves into principal directions of variation by a Principal Component Analysis (PCA) of the raw data or smoothed curves. PCA provides a simple recipe for dimension reduction by including only eigenvalue–eigenvector pairs where the estimated eigenvalue is relatively large. However, PCA is a second-order methodology in that it uses only the mean and covariance functions, and the Karhunen-Loève expansion based on its eigen-decomposition assumes a joint Gaussian distribution (Yao, Müller, and Wang, 2005).

The focus of current FDA methods on mean and covariance functions ignores subtler differences in distributions that could be better characterized by quantiles. Quantiles, like means, have tremendous data compression potential and are easy to interpret, a sine-qua-non in the century

of data. In some simulation experiments, we have shown that the quantiles of asymmetric data are estimated with substantial bias if the Gaussian model typical of FDA is assumed. Since these results were to be expected, we omit the details.

The paper is organized as follows. Section 2 introduces functional processes with heterogeneous shape characteristics and describes the modeling methodology. Section 3 develops the estimation procedure. Section 4 contains an extensive simulation study. Applications to the tractography data are presented in Section 5. A brief discussion is given in Section 6.

2 Functional models with spatially-varying marginal distributions

Let $\{Y_i(t_{ij}); t_{ij} \in \mathcal{T}\}$ with $j = 1, \ldots, m_i$ be the data for subject i, $i = 1, \ldots, N$. We assume that Y_i is a random curve defined on domain $\mathcal{T}$ and sampled at a grid of points $\{t_{i1}, \ldots, t_{im_i}\} \in \mathcal{T}$; typically we take $\mathcal{T} = [0, 1]$. Furthermore, suppose that

$$Y_i(t) = \mu(t) + \sigma(t) G^{-1}\{W_i(t); \alpha(t)\} \tag{1}$$

where $\mu(t)$ is the mean function, $\sigma(t)$ is the standard deviation function, and $W_i(t)$ is a latent process such that $W_i(t)$ is uniformly $(0, 1)$ distributed for each t. Here $G(\cdot; \alpha)$ is a cumulative distribution function in a parametric family of distributions with zero mean, unit variance, and shape parameter vector α. For example, α can be the scalar shape parameter of the skew normal distribution (Azzalini, 1985) or the two-dimensional shape parameter vector of the skew t distribution due to Azzalini and Capitanio (2003) or a somewhat different skew t distribution of Fernandez and Steel (1998). Also, $G^{-1}(\cdot; \alpha)$ denotes the usual inverse CDF. We assume that the shape parameter $\alpha(t)$ varies smoothly with t. To simplify notation, we sometimes use $G_t := G\{\cdot; \alpha(t)\}$. Our main objectives are to estimate the population functions $\mu(t)$, $\sigma(t)$, and $\alpha(t)$ and describe the dependence of the functional process.

We refer to model (1) as the quantile-induced functional model, because model (1) implies the following model for the pth quantile: $Q_p(t) = \mu(t) + \sigma(t) G_t^{-1}(p)$, $0 < p < 1$. The advantage of representing the functional data using model (1) is that it allows the specification of the dependence of the random process Y_i through the latent process W_i. Since $W_i(t)$ is uniformly distributed for each t, the joint distribution of $\{W_i(t_{i1}), \ldots, W_i(t_{im_i})\}$ is a copula, which we will model parametrically using Gaussian or t copulas.

2.1 Population-level parameters

We model the mean function $\mu(t)$, variance function $\sigma^2(t)$, and the shape parameter functions $\alpha(t)$ using penalized splines, though other basis expansion approximations can be used. Because the variance is positive, we use log-splines to model $\sigma^2(t)$. Let $\mathbf{B}(t) = \{B_1(t), \ldots B_q(t)\}^T$ be a truncated power series spline basis with dimension q having a fixed number of knots. Because the penalty prevents overfitting, the number of knots has very little effect on a penalized spline fit, provided the number of knots is sufficiently large to accommodate fine-level features in the data (Ruppert, 2002; Li and Ruppert 2008). However, an excess of knots will slow the computations. Therefore, in practice, we allow a different number of knots for μ, σ, and α, since some of these functions may require more knots than others. However, to keep notation simple, in our exposition we will use the same spline basis for each of these three functions. We write $\mu(t) = \mathbf{B}(t)^T\beta_\mu$ and $\log\{\sigma^2(t)\} = \mathbf{B}(t)^T\beta_\sigma$, where the spline bases and coefficient vectors are q-dimensional. In some cases it is desirable to model, not $\alpha(t)$, but a suitable transformation of $\alpha(t)$, say $h\{\alpha(t)\}$, where $h(\cdot)$ is a one-to-one transformation. For example, if any component of $\alpha(t)$ is positive we could log-transform that component. Our model for $\alpha(t)$ is $\alpha(t) = h^{-1}\{\mathbf{B}(t)^T\beta_\alpha\}$.

2.2 Functional dependence and copulas

A copula is a multivariate distribution function whose univariate marginal distributions are all uniform on (0,1). If $\mathbf{X} = (X_1, \ldots, X_n)$ is a random vector and if F_i is the continuous marginal distribution of X_i, then the distribution of $\{F_1(X_1), \ldots, F_n(X_n)\}$ is called the copula of $\mathbf{X}$. Copulas offer a convenient way of modeling multivariate observations because the modeling is broken into two independent parts: (1) modeling the dependencies through a copula; and (2) modeling the univariate marginal distributions (Sklar, 1959).

A Gaussian copula is the copula of some multivariate Gaussian distribution. The copula of a random vector is unchanged by strictly increasing transformations of its components, so a Gaussian copula is completely specified by a correlation matrix. If $\mathbf{X}$ has a Gaussian copula, this does not imply that $\mathbf{X}$ is Gaussian, only that $\mathbf{X}$ has the same copula as some Gaussian random vector. Stated differently, $\mathbf{X}$ has the same dependence structure as some multivariate normal random vector, but its marginal distributions need not be Gaussian. Similarly, a t-copula is the copula of

some multivariate t-distribution and is specified by a correlation matrix and a scalar degrees of freedom parameter.

Gaussian copulas offer a convenient way to model multivariate dependencies, because these are determined by a familiar quantity, a correlation matrix. The limitation of the Gaussian copula concerns the tail dependence behavior. The coefficient of the tail dependence between a pair (X, Y) of random variables is $\lim_{p\to\infty} P\{Y > F_Y^{-1}(p)|X > F_X^{-1}(p)\}$, which is easily shown to be symmetric in X and Y. For a bivariate Gaussian pair, this coefficient is zero unless their correlation equals 1 (Embrechts, McNeil & Straumann, 2002). This means that under a Gaussian copula, the components behave in the extreme tails as they were either independent, or perfectly correlated. In contrast, t-copulas exhibit tail dependence even in the case of zero-correlation; see McNeil, Frey, and Embrechts (2005). We found evidence of strong tail dependence in the DTI data; see Section 5.

3 Estimation Methodology

We use penalized maximum likelihood to estimate the mean, standard deviation and shape parameter functions, $\mu(t)$, $\sigma(t)$ and $\alpha(t)$. Each of these functions contains a reasonably large number of parameters and the correlation matrix of the Gaussian or t-copula is also of sizable dimension. Because of the large number of parameters, simultaneous estimation of all parameters can take an excessive amount of computational time. To circumvent this problem, we have developed a more rapid multi-stage estimation procedure.

3.1 Mean, variance, and shape parameter functions

To speed computations, the estimates of $\mu(t)$, $\sigma(t)$ and $\alpha(t)$ are obtained in two steps. In the first, the sample of curves is reduced to undersmoothed estimates of these three functions. In the second, these curve estimates are further smoothed by penalized splines.

Step 1: For simplicity, we start with the case where the functions are sampled on a common dense grid of points, so that $t_{ij} = t_j$ for $j = 1, \ldots, m$ and all i. Initial estimates of the mean, variance and shape parameter functions are constructed as follows: for each fixed j, define $\mu_j = \mu(t_j)$, $\sigma_j = \sigma(t_j)$ and $\alpha_j = \alpha(t_j)$. Then, the parametric estimates of μ_j, σ_j and α_j are obtained by

maximizing the pointwise likelihood function

$$\ell_j(\mu_j, \sigma_j, \alpha_j) = \sum_{i=1}^{n} \log[g\{(Y_{ij} - \mu_j)/\sigma_j; \alpha_j\}] - \log(\sigma_j), \tag{2}$$

where $g(x; \alpha_j) = \partial G(x; \alpha_j)/\partial x$ is the density function corresponding to the distribution G_{t_j}. For example, when G_t is in the skew normal or skew t family of distributions, the functions `sn.mle` or `st.mle` of the `sn` package for `R` (Azzalini, 2010) can compute these estimates.

Maximum likelihood estimation of α often requires some care. For example, if g is a skew normal density or a skew t density with known degrees of freedom, then the estimate of α is infinite with positive probability for moderate sample sizes (Azzalini, 1985; Genton, 2004; Sartori, 2006). This instability problem is due to the choice of parameterization in that large changes in α lead to small changes in the density. It should be noted, however, that infinite values of α give valid densities, the half-normal (or half t) densities. The problem is that some estimation methods, for example, spline smoothing, have problems accommodating infinite values. A possible remedy for this problem is mentioned in "*Alternative 1*" below.

When the curves are sparsely sampled, another estimation method is needed. Note that the estimates of μ_j, σ_j and α_j in (2) are obtained by local polynomial estimating equations (Carroll, Ruppert, and Welsh, 1998), with degree zero polynomials and a bandwidth so small that only data exactly at t_j are used to estimate the functions at t_j. For sparse and irregularly spaced data, local estimating equations with larger bandwidths and higher polynomial degree can be used to compute estimates on a grid $t_1, \ldots, t_m$, say. The estimates should be undersmoothed, since they will be smoothed further in Step 2.

In summary, regardless of whether we have sparse or dense data, at the end of Step 1 we have undersmoothed estimates on a grid $t_1, \ldots, t_m$. These estimates will be denoted by $\widetilde{\mu}_j$, $\widetilde{\sigma}_j$, and $\widetilde{\alpha}_j$, $j = 1, \ldots, m$. Step 2 uses $\widetilde{\mu}_j$ and $\widetilde{\sigma}_j$ and, in *Alternative 1*, $\widetilde{\alpha}_j$ is also used.

Step 2: The final estimates of $\mu(t)$, $\sigma(t)$ and $\alpha(t)$ are constructed by smoothing the Step 1 estimates using penalized splines. The mean parameter β_μ, where $\mu(t) = \mathbf{B}(t)^T\beta_\mu$, is estimated by minimizing the penalty criterion $PL_\mu(\beta_\mu) = \sum_{j=1}^{m}\left(\widetilde{\mu}_j - \mathbf{B}_j^T\beta_\mu\right)^2 + \lambda_\mu\Omega_\mu(\beta_\mu)$, where $\Omega_\mu(\beta_\mu) = \beta_\mu^T D_\mu \beta_\mu$, $\mathbf{B}_j = \mathbf{B}(t_j)$, and D_μ is a $q \times q$ penalty matrix; see Ruppert et al. (2003) for a discussion of penalty matrices. In a similar way $\widehat{\beta}_\sigma$ is obtained by minimizing $PL_\sigma(\beta_\sigma) = \sum_{j=1}^{m}\left\{2\log(\widetilde{\sigma}_j) - \mathbf{B}_j^T\beta_\sigma\right\}^2 + \lambda_\sigma\Omega_\sigma(\beta_\sigma)$ for $\Omega_\sigma(\beta_\sigma) = \beta_\sigma^T D_\sigma \beta_\sigma$.

For the shape function parameters, we propose a penalized marginal likelihood criterion where the mean and variance functions parameters are fixed at estimates (see also Sartori, 2006). More specifically, let $\widehat{Y}_{ij} = \{Y_{ij} - \widehat{\mu}(t_{ij})\}/\widehat{\sigma}(t_{ij})$ be the standardized observations, using the estimates of the mean $\widehat{\mu}(\cdot)$ and variance functions $\widehat{\sigma}(\cdot)$ obtained above. Define the penalized marginal likelihood function:

$$PL_\alpha(\beta_\alpha) = -2\sum_{i=1}^{n}\sum_{j=1}^{m}\left\{\ell_{ij}(\beta_\alpha; \widehat{Y}_{ij})\right\} + \lambda_\alpha \Omega_\alpha(\beta_\alpha),$$

where $\ell_{ij}(\beta_\alpha; \widehat{Y}_{ij}) = \log\left[g\left\{\widehat{Y}_{ij}; h^{-1}\left(\mathbf{B}_j^T\beta_\alpha\right)\right\}\right]$ is, up to an additive constant, the log-likelihood function corresponding to the distribution of $\widehat{Y}_i(t_{ij})$. We use $\Omega_\alpha(\beta_\alpha) = \beta_\alpha^T D_\alpha \beta_\alpha$ as the roughness penalty, where D_α is the roughness penalty matrix. The shape parameter β_α is obtained as the minimizer of the penalized criterion $PL_\alpha(\beta_\alpha)$.

For the mean and variance estimation, we selected the smoothing parameters λ_μ and λ_σ using the restricted maximum likelihood estimation (REML) (Wood, 2006). For the shape parameter function estimation λ_α we used using the corrected AIC criterion (Ruppert et al., 2003); other choices, such as CSV or AIC can be used. The function `optim` of the `stats` package for `R` with the BFGS algorithm can be used for optimization. Let $\widehat{\mu}(t) = \mathbf{B}(t)^T\widehat{\beta}_\mu$, $\widehat{\sigma}(t) = \exp\{\mathbf{B}(t)^T\widehat{\beta}_\sigma/2\}$, and $\widehat{\alpha}(t) = h^{-1}\left\{\mathbf{B}(t)^T\widehat{\beta}_\alpha\right\}$ be the estimated mean, standard deviation and shape parameter functions respectively.

Alternative 1. One fast alternative is to estimate the shape parameter functions using the same approach as for the mean and variance function. To handle infinite values of α_j, we model transformed parameters $h(\alpha_j)$ instead. For example, if $G(\cdot;\alpha)$, $-\infty \leq \alpha \leq \infty$, is the skew normal distribution, then an appropriate transformation is $h(\cdot) = \Phi(\cdot/s)$, the normal CDF with standard deviation s, which maps $[-\infty, \infty]$ to $[0, 1]$; the choice $s = 5$ was used in our simulation study and worked satisfactorily. This transformation eliminates the infinite values that cause problems. The shape parameter β_α minimizes the penalized criterion $PL_\alpha(\beta_\alpha) = \sum_{j=1}^{m}\{h(\widetilde{\alpha}_j) - \mathbf{B}_j^T\beta_\alpha\}^2 + \lambda\beta_\alpha^T D_\alpha \beta_\alpha$, with λ and D_α defined as above. While this approach is extremely fast, it estimates the shape parameter function with higher variability.

Alternative 2. The estimation method described in this section estimates $\mu(t)$, $\sigma(t)$ and $\alpha(t)$ without accounting for the dependence of the process. One way to incorporate the functional dependence is via a penalized full likelihood approach. We briefly outline this idea. Let $C(;\Omega)$

be a copula distribution, where Ω is the set of copula parameters. Estimation of Ω is discussed in Section 3.2. Denote by $c(\cdot;\Omega)$ the copula density. It follows that the log-likelihood of the full model can be written as:

$$\ell(\beta_\mu,\beta_\sigma,\beta_\alpha,\Omega)=\sum_{i=1}^{N}\sum_{j=1}^{m}\ell_{ij}(\beta_\mu,\beta_\sigma,\beta_\alpha;Y_{ij})+\sum_{i=1}^{N}\log\left[c_{\mathbf{t_i}}\{F_{Y,t_{i1}}(Y_{i1}),\ldots,F_{Y,t_{im}}(Y_{im});\Omega\}\right], \qquad (3)$$

where $\mathbf{t_i}=(t_{i1},\ldots,t_{im})$, $Y_{ij}=Y_i(t_{ij})$, $\ell_{ij}(\beta_\mu,\beta_\sigma,\beta_\alpha;Y_{ij})=\log\Big[g\Big\{(Y_{ij}-\mathbf{B}_{ij}^T\beta_\mu)/\exp(\mathbf{B}_{ij}^T\beta_\sigma/2);\ h^{-1}\left(\mathbf{B}_{ij}^T\beta_\alpha\right)\Big\}\Big]-\mathbf{B}_{ij}^T\beta_\sigma/2$ is the penalized log-likelihood function of Y_{ij} and $F_{Y,t_{ij}}(Y_{ij})=G\Big\{(Y_{ij}-\mathbf{B}_{ij}^T\beta_\mu)/\exp(\mathbf{B}_{ij}^T\beta_\sigma/2);\ h^{-1}\left(\mathbf{B}_{ij}^T\beta_\alpha\right)\Big\}$. Here $\mathbf{B}_{ij}=\mathbf{B}(t_{ij})$ for all i and j. To avoid over-smoothness of the population-level functions, which may be caused by large-dimensional function bases, a penalized criterion will be used instead. The mean, variance and shape parameter functions parameters β_μ,β_σ and β_α are chosen to minimize the penalized log-likelihood $PL(\beta_\mu,\beta_\sigma,\beta_\alpha,\Omega)=-2N^{-1}\ell(\beta_\mu,\beta_\sigma,\beta_\alpha,\Omega)+\lambda_\mu\Omega_\mu(\beta_\mu)+\lambda_\sigma\Omega_\sigma(\beta_\sigma)+\lambda_\alpha\Omega_\alpha(\beta_\alpha)$, where Ω_μ, Ω_σ and Ω_α are roughness penalties.

Our experience from simulation studies is that *Alternative 2* estimates of the population functions $\mu(\cdot)$, $\sigma(\cdot)$, and $\alpha(\cdot)$ are not very different from the estimates that use the working independence assumption, while the computational cost of *Alternative 2* is considerably higher.

3.2 Calibrating copulas

Estimation of the copula parameters is done somewhat differently, according to the type of copula used. A Gaussian copula requires only estimation of the correlation matrix of the underlying Gaussian distribution. A t-copula requires estimation of a correlation matrix and a degree of freedom parameter. First, we describe estimation of the correlation matrix, which is similar for the two copula families, and then describe estimation of the degree of freedom parameter.

If a Gaussian copula is assumed, then the copula correlation is precisely the Pearson correlation of the Gaussian process

$$Z_i(t)=\Phi^{-1}\Big(G\Big[\{Y_i(t)-\mu(t)\}/\sigma(t);\alpha(t)\Big]\Big). \qquad (4)$$

This process has standard normal marginal distributions, so $K(s,t)=\mathrm{cov}\{Z_i(s),Z_i(t)\}$ is both the process covariance and correlation function, and thus it is the latent copula correlation function. One straightforward way to estimate K is to use method of moment estimators of the covariance

function corresponding to the approximately Gaussian processes obtained by replacing the population functions by their estimates, $\widehat{Z}_i(t_{ij}) = \Phi^{-1}\Big(G\Big[\{Y_i(t_{ij}) - \widehat{\mu}(t_{ij})\}/\widehat{\sigma}(t_{ij}); \widehat{\alpha}(t_{ij})\Big]\Big)$. We model $\widehat{Z}_i(t)$ as the sum of two independent components: a finite basis expansion process, say $\widetilde{Z}_i(t)$ with covariance function $\widetilde{G}(s,t)$, and a white noise process with variance $\sigma^2_\epsilon 1(t=s)$. By a finite basis expansion process, we mean a Karhunen-Loève expansion truncated at a finite, and generally small, number of eigenvectors. This is a relatively common technique for handling the high correlation in functional data; see for example Kneip and Sarda (2010).

Let $\widehat{G}(s,t)$ be the method of moment estimator of $\mathrm{cov}\{\widehat{Z}_i(s), \widehat{Z}_i(t)\}$. Let $\widetilde{G}(s,t) = \widehat{G}(s,t)$ for $s \neq t$; for $s = t$, $\widetilde{G}(t,t)$ is estimated using a bivariate thin-plate spline smoother applied to $\widehat{G}(t,s)$ for $t \neq s$. This approach to estimation of the diagonal elements removes the "nugget effect" due to the white noise term, and was proposed by Staniswalis and Lee (1998). Because $\widetilde{Z}_i$ is modeled by a finite basis expansion process, it makes sense to use a reduced rank approximation of the covariance of $\widetilde{Z}_i$. We estimate σ^2_ϵ by $\widehat{\sigma}^2_\epsilon = \int_0^1 \{\widehat{G}(t,t) - \widetilde{G}(t,t)\}\, dt$; if this estimate is not positive, then it is replaced by a small positive number. Assume that the reduced rank approximation of $\widetilde{G}$ is $\Psi\Lambda_L\Psi^T$, where Λ_L is diagonal matrix of dimension $L < m$ and Ψ is an $m \times L$ matrix with orthogonal columns. The correlation matrix of the copula density is the correlation matrix corresponding to the covariance matrix $\widetilde{G} + \widehat{\sigma}^2_\epsilon I_m$. Because $\sigma^2_\epsilon > 0$, this correlation matrix is guaranteed to be positive definite.

The estimates of the correlation matrix just described use the estimates of the population mean, variance, and shape functions, and will be degraded by errors in the estimation of these functions. As an alternative, the sample Kendall's tau matrix can be used to estimate the latent correlation matrix K (McNeil, et al., 2005, section 5.3.2). Estimation of the correlation matrix using Kendall's tau is easiest when $t_{ij} = t_j$ for all i, which is the case we will discuss. The sample Kendall's tau is invariant to increasing transformations, so that using the data $Y_i(t_j)$, or the transformed $\widehat{Z}_i(t_j)$ gives the same result. Using (5.32) of McNeil, et al. (2005), let $\widehat{\rho}_\tau(t_j, t_{j'})$ be the sample Kendall's tau between $\{Y_1(t_j), \ldots, Y_N(t_j)\}$ and $\{Y_1(t_{j'}), \ldots, Y_N(t_{j'})\}$ and define $\widehat{G}(t_j, t_{j'}) = \sin\{\pi\widehat{\rho}_\tau(t_j, t_{j'})/2\}$. Then, the estimation of K proceeds as described above.

In our extensive simulation study, the two methods of estimating the copula correlation matrix performed almost identically. When the data are sparse and $t_{ij} = t_j$ for all i does not hold, then one can estimate the correlation matrix using the techniques in Yao et al. (2005). A full investigation

of the sparse case is beyond the scope of this paper.

Suppose now that a t-copula is assumed. Consider the transformed random functions

$$Z_i^*(t) = T_\nu^{-1}\Big(G\Big[\{Y_i(t) - \mu(t)\}/\sigma(t); \alpha(t)\Big]\Big), \tag{5}$$

where T_ν is the t-CDF with ν degrees of freedom; these curves are distributed according to a t-process with ν degrees of freedom. One easy way to understand the t-process is through the following explicit construction. Let $\{Z_i(t) : t \in \mathcal{T}\}$ be a Gaussian process with mean 0 and covariance function $K(s,t) = \text{cov}\{Z_i(s), Z_i(t)\}$ such that $K(t,t) = 1$, so that $K(t,t)$ is also the correlation function. Write $Z_i^*(t) = Z_i(t)/\sqrt{\chi^2_{\nu,i}/\nu}$, where $\chi^2_{\nu,i}$ is a chi-squared random variable with ν degrees of freedom and independent of Z_i. If $\nu > 2$, then the covariance function of Z_i^* is $K_Z^*(s,t) = \nu/(\nu-2)K(s,t)$ and so its correlation function is $K(s,t)$. If $\nu \leq 2$, then Z_i^* has infinite second moments, so it does not have a covariance or a correlation function. However, K can still be regarded as an "association function" of Z_i^* and it is the correlation function of Z_i.

Calibrating the t-copula requires estimation of both K and ν. Estimation of K can be achieved using Kendall's tau as in the case of Gaussian copula. To estimate ν, we use a "pseudo-likelihood" method. Pseudo-likelihood means estimating certain parameters by maximizing the likelihood with all other parameters fixed at estimates. Denote by $\widehat{K}$ the estimate of the latent correlation matrix. Let $\widehat{F}_j$ be the empirical distribution function of $\{Y_i(t_j) : i = 1, \ldots, N\}$ except that the denominator is $N+1$, and not N. Let $\widehat{W}_{ij} = \widehat{F}_j\{Y_i(t_j)\} = R_{ij}/(N+1)$, where R_{ij} is the rank of $Y_i(t_j)$ among $\{Y_i(t_j) : i = 1, \ldots, N\}$, where $j = 1, \ldots m$. The "full pseudo-likelihood" method for estimation of ν is to fit the t-copula to the $\widehat{W}_{ij}$ with K fixed at the estimate $\widehat{K}$.

4 Simulation studies

We conducted an extensive simulation study to assess the performance of the proposed estimation procedures described in Section 3. In this section, we summarize the main findings based on 100 data sets, each consisting of $N = 200$ random trajectories generated from model (1). Each curve was sampled on a grid of equi-spaced timepoints $\{t_{ij} : j = 1, \ldots, 80\}$ in $[0,1]$, where $i = 1, \ldots, N$. The simulated processes use $W_i(t) = F_{X,t}\{X_i(t)\}$, where X_i's are a sample of mean-zero Gaussian curves and $F_{X,t}(\cdot)$ is the CDF of $X_i(t)$ for all $0 \leq t \leq 1$.

Two main covariance structures of the underlying processes are considered:

I. Finite basis expansion of the covariance function plus white noise. We have $X_i(t) = Z_i(t) + \epsilon_i(t)$, where $\epsilon_i(t)$ are assumed to be independent $N(0, \sigma^2 = 0.10)$ and Z_i's are a sample of random curves with mean 0 and covariance function $K_Z(s,t) = \text{cov}\{Z_i(t), Z_i(s)\}$. The covariance function is assumed to have the expansion $K_Z(s,t) = \sum_{\ell=1}^{L} \lambda_\ell \psi_\ell(s)\psi_\ell(t)$ in terms of eigenfunctions ψ_ℓ's and eigenvalues λ_ℓ's. In the simulations, the covariance function had $L = 3$ eigenfunctions:

(i) $\psi_1(t) = \sqrt{2}\sin(2\pi t)$, $\psi_2(t) = \sqrt{2}\cos(4\pi t)$ and $\psi_3(t) = \sqrt{2}\sin(4\pi t)$; or

(ii) $\psi_1(t) = \sqrt{3}(2t^2 - 1)$, $\psi_2(t) = \sqrt{5}(6t^2 - 6t + 1)$ and $\psi_3(t) = \sqrt{7}(20t^3 - 30t^2 + 12t - 1)$,

where $0 \le t \le 1$. We choose $\lambda_\ell = (1/2)^{\ell-1}$ for $\ell = 1, 2, 3$.

II. Matern covariance structure. We let X_i be Gaussian process with mean zero and Matérn auto-correlation function

$$\rho(\Delta; \phi, \kappa) = \frac{1}{2^{1-\kappa}\Gamma(\kappa)} \left(\frac{2\kappa^{1/2}\Delta}{\phi} \right)^{\kappa} K_\kappa \left(\frac{2\kappa^{1/2}\Delta}{\phi} \right) \tag{6}$$

with (i) range $\phi = 0.07$ and order $\kappa = 1$; and (ii) with range $\phi = 0.14$ and order $\kappa = 1$. For both cases the pointwise variance is set to 1. Here K_κ is the modified Bessel function of order κ.

We set G_t to be the standardized skew normal distribution with shape parameter $\alpha(t)$, which is implemented in the `sn` package of `R`. The standardization is such that the resulting distribution has mean 0 and variance 1. We consider all the possible combinations from the following scenarios:

1. mean function: (a) $\mu(t) = 6$; and (b) $\mu(t) = -2.2t^5 + 3t^3 - 1.2t + 0.7$;
2. variance function: (a) $\sigma^2(t) = \exp(-5)$; and (b) $\sigma^2(t) = \{2.2t^5 - 3t^3 + 1.2t + 0.3\}/\exp(4)$;
3. shape parameter function: (a) $\alpha(t) = 0$; (b) $\alpha(t) = -21(t \le 0.5) + 41(t > 0.5)$;

   (c) $\alpha(t) = 5t^2 - 19t + 5$; and (d) $\alpha(t) = -10\sin(2\pi t)$, for $t \in [0, 1]$.

For the estimation of the mean and variance functions, the smoothing parameters were selected using REML implemented in the `R` package `mgcv` (Wood, 2006). For the estimation of the shape parameter functions, we used cubic regression splines with 5 knots, with the smoothing parameter selected by the corrected AIC criterion. We used a Gaussian copula to model dependence. For bivariate smoothing of the covariance function, tensor product penalized cubic regression splines with 10 knots per dimension were used, and REML estimation was used for the selection of the smoothing parameter (Wood, 2006).

Figure 2 shows a data set generated from covariance structure (IIii), mean function (1b), vari-

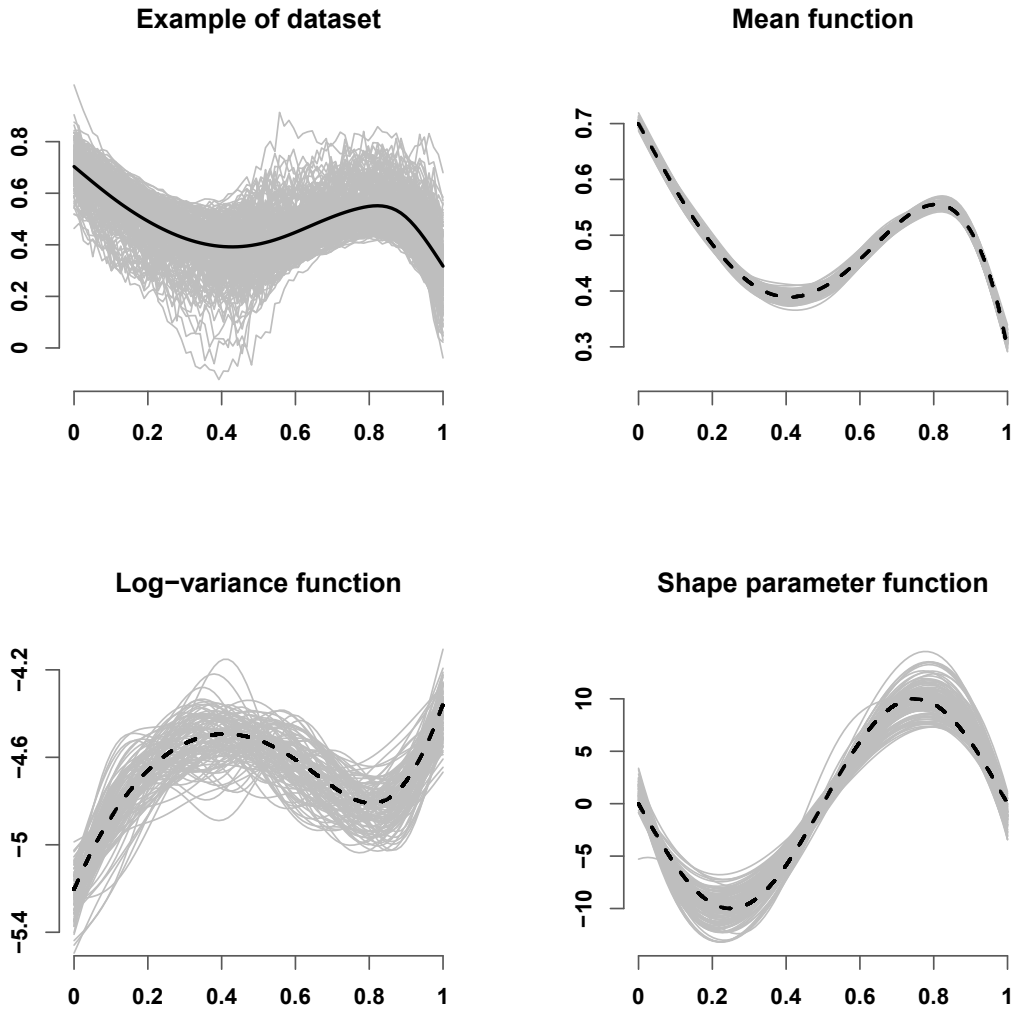

Figure 2: Simulated data. Displayed are: one dataset along with an estimate of the mean as a dashed line (top left panel), the 100 estimates of mean (top, right), log-variance (bottom, left) and shape parameter (bottom, right) functions.

ance function (2b), and shape function (3d) along with the estimates of the mean, log-variance and shape parameter functions from 100 simulated data sets. The estimates of all population-level functions show nearly no bias and various degrees of smoothness. In particular, the log-variance estimates are somewhat undersmoothed, which is what is expected in the case of smoothing dependent data with a working independence assumption. Nonetheless, the undersmoothing does not degrade the MSE by much, as will be seen soon. In Table 1, we report estimates of the squared root of the integrated mean squared error (IMSE), the integrated squared bias (IBIAS), and the integrated variance (IVAR). For example, for $\widehat{\mu}(t)$, these quantities are defined as: IMSE $= \int_0^1 E[\{\widehat{\mu}(t) - \mu(t)\}^2]\, dt$, IVAR $= \int_0^1 \mathrm{Var}\{\widehat{\mu}(t)\}\, dt$ and IBIAS $= \int_0^1 [E\{\widehat{\mu}(t)\} - \mu(t)]^2\, dt$. Here $E\{\widehat{\mu}(t)\}$ and $\mathrm{Var}\{\widehat{\mu}(t)\}$ denote the sample mean and sample variance of $\widehat{\mu}(t)$, for $0 \leq t \leq 1$. The results confirm our previous observations: the mean function estimator has very small IMSE, irrespective of the covariance structure. The somewhat larger IMSE of the log-variance function estimator is caused by its larger variability. For the shape parameter function estimator both the bias and the variability cause a larger IMSE. In general, estimation of the mean, log-variance and shape functions is not affected by the structure of the dependence of the latent process in cases where the latent process exhibits stronger dependence. As expected, when the latent process has very weak dependence, such as in scenario (II i), better estimates are obtained for all the population functions.

Figure 3 illustrates the performance of our methodology with respect to capturing the true dependence of the functional process. Figure 3 displays the estimates of the first three eigenfunctions and eigenvalues of the copula correlation matrix in 100 simulated data sets from scenario (Ii). The true eigenfunctions and eigenvalues correspond to the correlation function derived from the model specified by the covariance structure (Ii); these eigenfunctions/eigenvalues are different from the ones defying the covariance functions. The copula correlation was estimated using Kendall's tau; however, similar performance was noted when the sample Pearson correlation of the transformed curves was used.

One important advantage of our model is that it provides access to explicit estimates of the pointwise quantile functions. Let $p \in (0, 1)$ be some specified quantile level; the estimated quantile function is $\widehat{Q}_p(t) = \widehat{\mu}(t) + \widehat{\sigma}(t) G^{-1}\{p; \widehat{\alpha}(t)\}$. Table 2 shows the estimates of the IMSE of the estimated quantile functions $\widehat{Q}_p(t)$ for several levels p, for one particular scenario. Our proposed

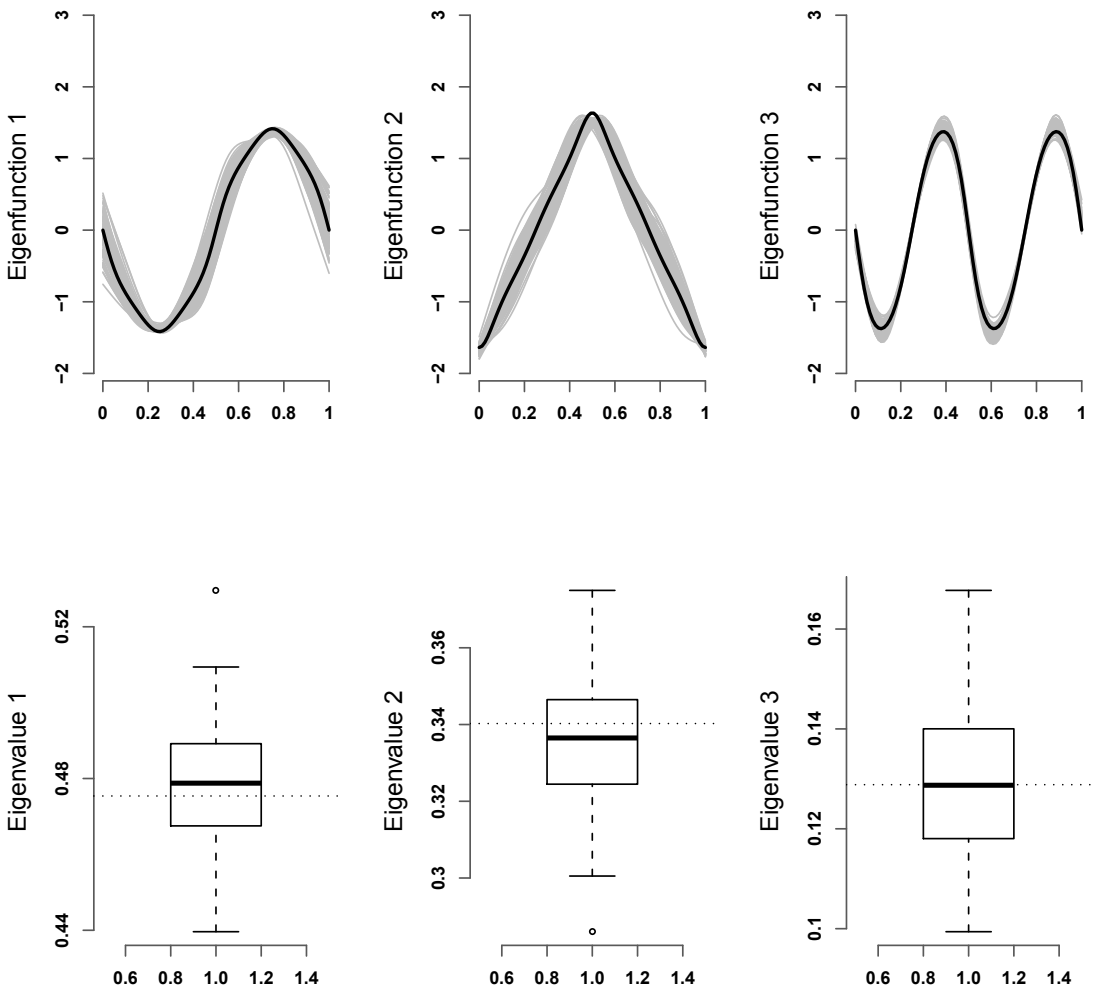

Figure 3: Estimated eigenfunctions (gray lines, top panels) and eigenvalues (bottom panels) and their corresponding true eigenfunctions (black lines) / eigenvalues (dotted line) derived from the true model, when the model uses mean function (1b), variance function (2b), shape function (3d), and covariance structure (Ii).

Table 1: Estimates of the squared root of IBIAS, the root of IVAR and the root of IMSE for the estimates of the mean function (1b), the log-variance function (2b), the shape parameter function (3d) when the latent process is specified as in (Ii) and (Iii) or in (IIi) and (IIii).

Latent Covariance	Population Level Functions	$\sqrt{IBIAS}$ ($\times 10^2$)	$\sqrt{IVAR}$ ($\times 10^2$)	$\sqrt{IMSE}$ ($\times 10^2$)
(I i)	Mean	0.27	0.68	0.72
(I ii)	Mean	0.26	0.71	0.76
(II i)	Mean	0.27	0.57	0.63
(II ii)	Mean	0.27	0.66	0.71
(I i)	Log-variance	1.94	10.18	10.32
(I ii)	Log-variance	1.90	10.74	10.86
(II i)	Log-variance	1.10	7.72	7.76
(II ii)	Log-variance	1.31	9.57	9.62
(I i)	Shape	66.76	117.13	134.31
(I ii)	Shape	64.34	106.59	124.05
(II i)	Shape	61.62	63.51	88.26
(II ii)	Shape	65.68	79.58	102.87

method gives estimated quantile functions with a better integrated mean square error and integrated variance, though with a slightly larger integrated bias, compared to the counterparts given by the pointwise sample quantiles. However, the estimated quantile functions are smooth functions, as opposed to the sample quantile functions which are not; see Figure 4 which plots the estimated and empirical quantile functions of levels 1%, 10% and 50%.

5 Application: DTI tractography

We apply our proposed method to a study of the diffusion characteristics of white matter tracts in patients with multiple sclerosis (MS) and healthy controls. White matter tracts consist of axons, often-long processes that connect one nerve cell to another and that convey electrical information in the form of nerve impulses. These axons are covered with a white fatty coating, called myelin. The myelin sheath helps the nerve transmit signals at a very fast rate. Myelin damage, as seen in MS and other demyelinating diseases, impairs axonal conduction and can be associated with axonal degeneration. Inflammatory demyelination and axon damage in the corpus callosum tract

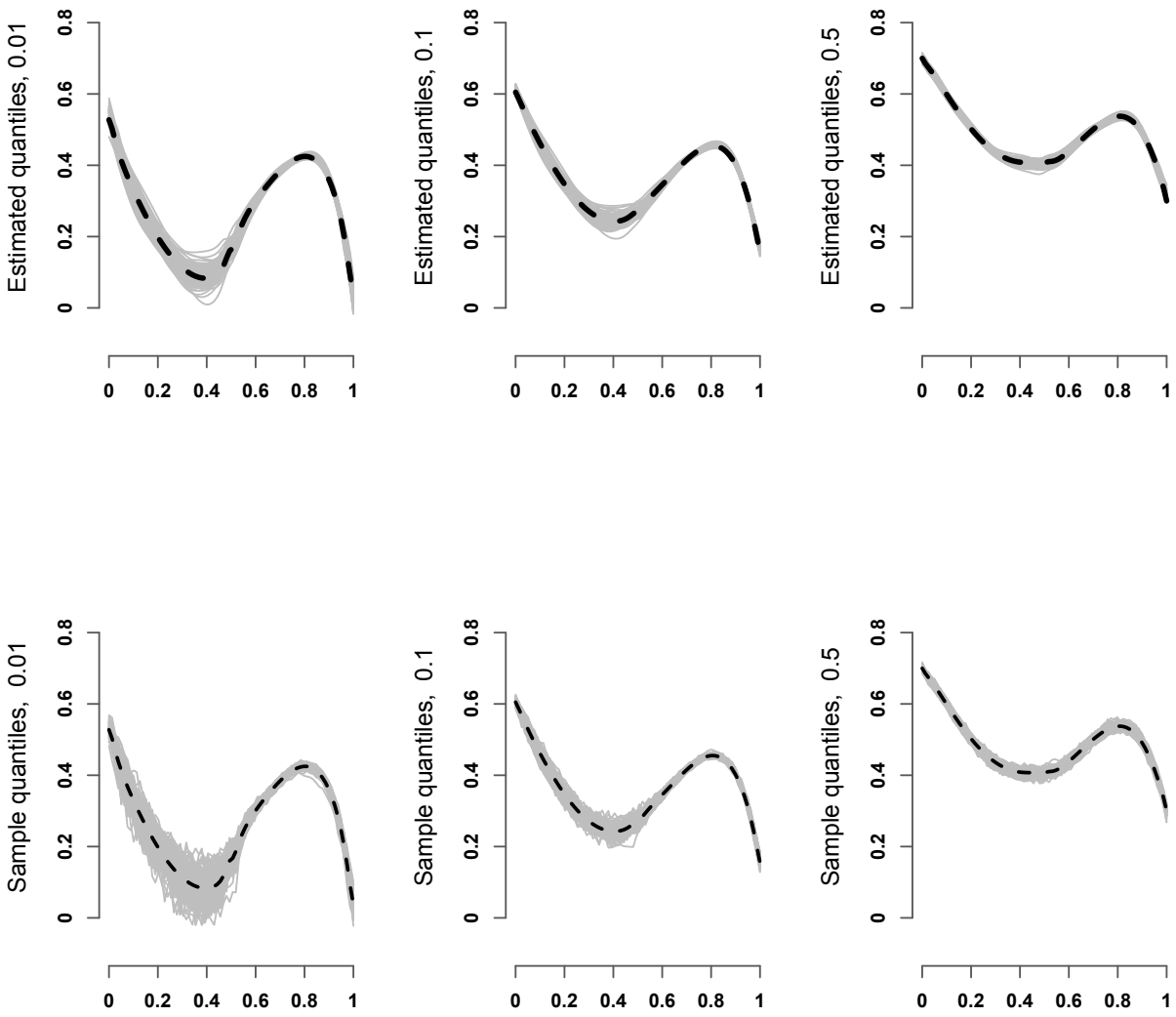

Figure 4: True and estimated quantiles when the simulation model uses mean function (1b), variance function (2b), shape function (3d), and latent covariance structure (Ii). Shown are the model-based estimated pointwise quantile function (top panels), the pointwise empirical quantiles (bottom panels) for each of 100 simulations and the true quantile functions (dashed) for levels $p = 1\%, 10\%$ and 50%.

Table 2: Estimates of the squared root of IBIAS, the root of IVAR and the root of IMSE for the estimates of the quantile functions of levels 1%, 2%, 5% and 10%, for the case when the true mean function is (1b), the log-variance function is (2b), the shape parameter function is (3d) and the latent process is specified by (Ii).

Method	Quantile level	$\sqrt{IBIAS}(\times 10^2)$	$\sqrt{IVAR}(\times 10^2)$	$\sqrt{IMSE}(\times 10^2)$
Model based	1%	0.43	1.58	1.63
Empirical	1%	0.32	2.67	2.68
Model based	2%	0.37	1.40	1.44
Empirical	2%	0.31	2.07	2.08
Model based	5%	0.31	1.16	1.19
Empirical	5%	0.20	1.49	1.50
Model based	10%	0.29	0.97	1.01
Empirical	10%	0.11	1.19	1.19
Model based	50%	0.30	0.64	0.71
Empirical	50%	0.06	0.89	0.89

are prominent features of MS and may partially account for impaired performance on complex tasks (Ozturk et al., 2010).

DTI reveals microscopic details about the architecture of the white matter tracts by measuring the three-dimensional directions of water diffusion in the brain (Basser et al. 1994, 2000). Our study involves measurements of the parallel diffusivity within the corpus callosum for 162 MS patients and 42 healthy controls. Goldsmith et al. (2010) used parallel diffusivity within left intracranial cortico-spinal tracts to classify subjects as MS cases or controls. Figure 1 displays tract profiles sampled at 93 locations in MS patients and controls.

A quick inspection of Figure 1 suggests that the controls and MS patients have somewhat similar means but different variability and asymmetry across tract location. We assumed that at each tract location, the parallel diffusivity for each group has a skew t distribution with constant, but unknown, degrees of freedom and location-specific skewness. The degrees of freedom parameter (ν) was estimated through maximum likelihood under a working independence assumption. This approach yields $\widehat{\nu} = 6.79$ for the MS group and $\widehat{\nu} = 16482.29$ for the control group. Because $\widehat{\nu}$

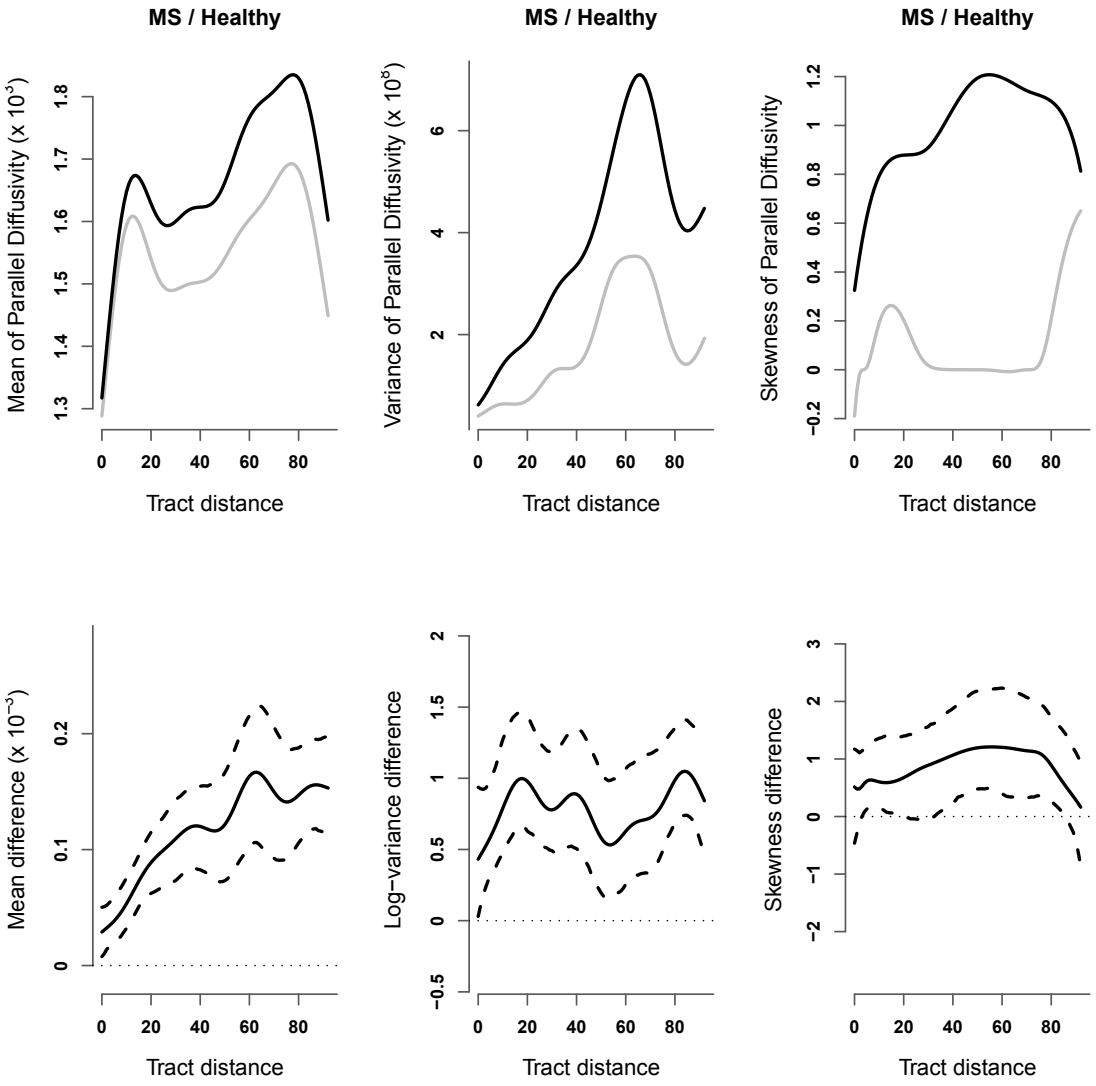

Figure 5: Estimated mean, variance, and third-moment skewness coefficient functions of the parallel diffusivity profiles within the corpus callosum tracts of MS patients and controls. Top panel: estimates for the MS group (black lines) and control group (grey lines); the skewness coefficient functions were calculated using the estimates of the shape parameter function and, for the t-copula used for the MS cases, the degrees of freedom. Bottom panel: Estimated differences between the mean (left), log-variance (middle), and skewness (right) functions of the MS and control groups, with 90% confidence intervals using 1000 bootstrap samples (dashed).

of the control group is very large, and since the skew t distribution reduces to skew normal when $\nu = \infty$, we used the skew normal for the control group.

In the top left panel of Figure 5, the estimated mean of parallel diffusivity is slightly larger for cases than for controls. Figure 5 (bottom left panel) shows the 90% bootstrap confidence intervals (based on 1000 resamples) for the differences between the means of the groups and indicates that the observed differences are statistically significant. The middle panels in Figure 5 depict significantly higher variability for the parallel diffusivity in the MS group than in the control group. The rightmost panels study the third-moment skewness coefficient functions of the two groups. For the control group, the skewness function depends solely on the shape parameter that governs a skew normal distribution; while it depends on both the degrees of freedom and the shape parameters for the MS group. The figure shows that the skewness function in the MS sample of parallel diffusivity is significantly larger than the skewness function in the control sample for locations 30-85 along the corpus callosum tract.

Figure 6 illustrates the comparison between the quantile functions of the parallel diffusivity profiles corresponding to the two groups at the quantile levels 1%, 50%, 95% and 99%. The lower quantile functions seem comparable for the two groups: this is expected when the mean functions are close to each other and the variance functions are small, which is our case. The higher level quantiles are indicative of the differences between the two groups in both the shape parameter and the variance. Figure 6 (bottom panel) shows the 90% pointwise confidence intervals of the estimated difference between the quantile functions of the two groups: the quantile functions for the MS group are significantly larger than the corresponding ones for the control group for probabilities of 50% and higher.

We first used Gaussian copulas, with group specific parameters, to model the dependencies in the parallel diffusivity profiles. Figure 7 shows the estimated eigenfunctions and eigenvalues of the correlation matrices for the MS patients and control groups. The similarity between these estimates in the two groups is remarkable. It suggests that to compare groups of curves, models with common correlation functions can be used in future analyses.

In addition, t-copulas were fit to the MS and control groups. The correlation matrices, estimated by the Kendall's tau approach in Section 3.2, are very similar with to the Gaussian copula correlation matrices. The maximum likelihood estimates of copula degrees of freedom were 3.30

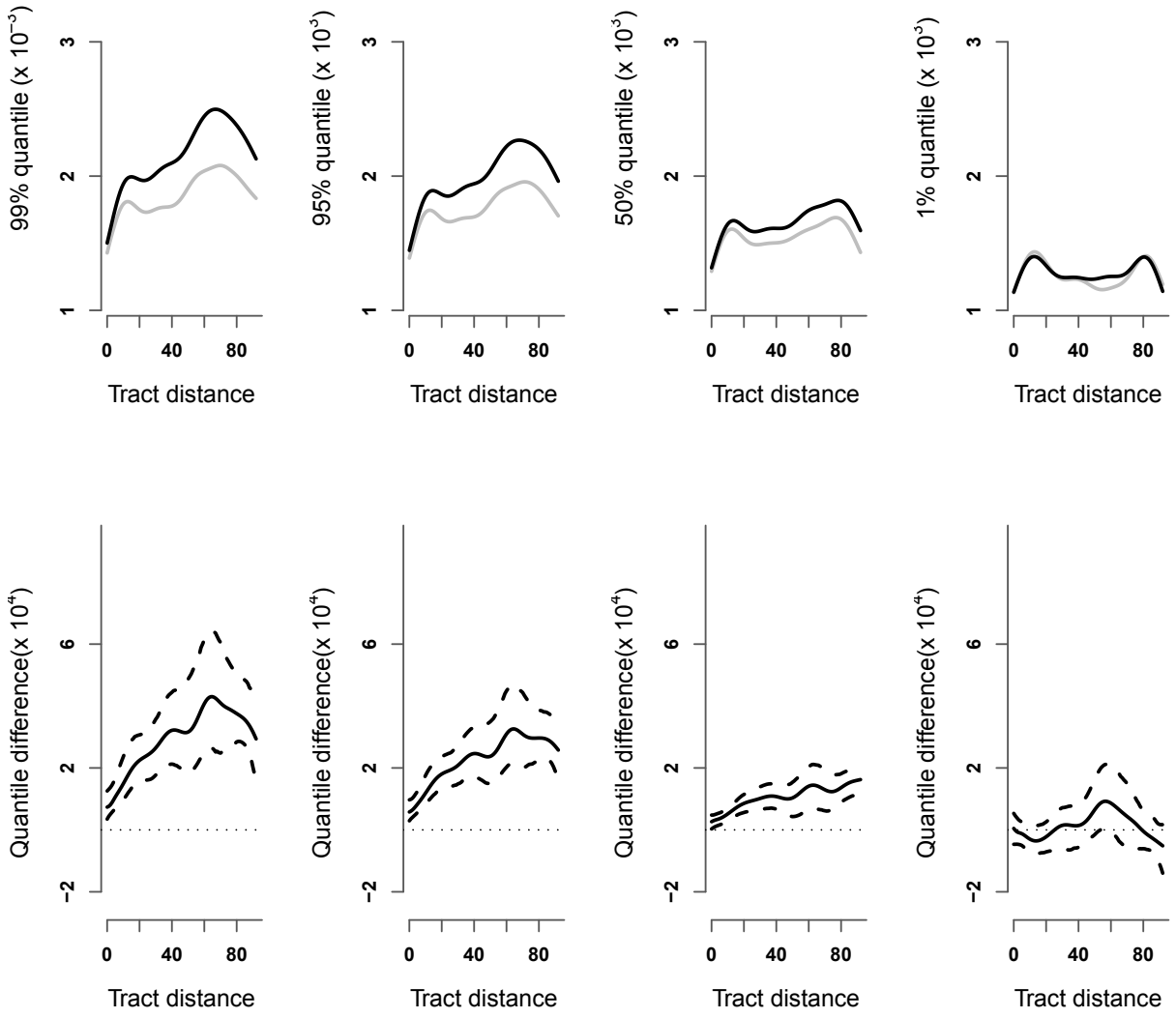

Figure 6: Estimated quantiles of the parallel diffusivity profiles within the corpus callosum tracts of the MS patients and controls. Top panel: Estimated quantiles functions for the MS group (black lines) and the control group (grey lines). Bottom panel: Estimated difference between the quantiles of the MS and control groups (solid lines) with 90% pointwise confidence intervals (dashed lines) from 1000 bootstrap replicates.

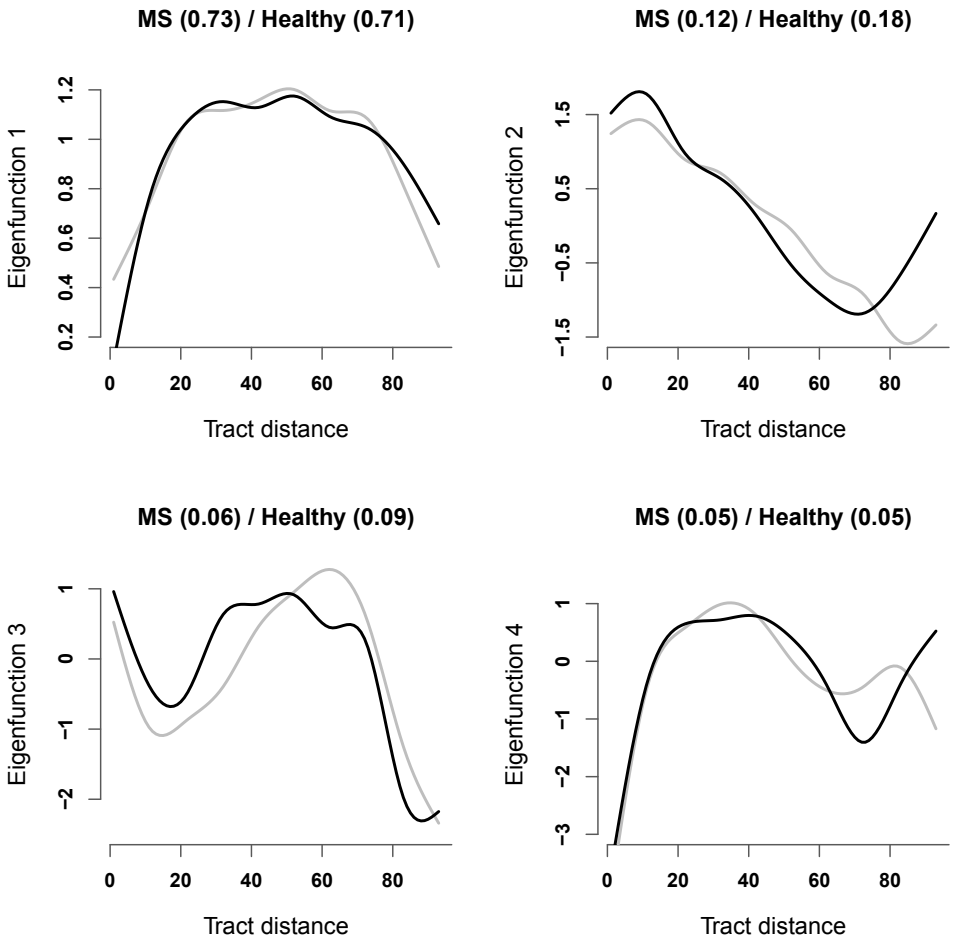

Figure 7: Estimated eigenfunctions of the latent Gaussian copula correlation functions in the MS (black) and control (grey) groups. Percentages of explained variance are given in parenthesis corresponding to MS and healthy subjects.

for the control group and 4.88 for the MS group. (These estimates should not be confused with the estimates of ν for the marginal skew t distributions.) Such small values of the copula degrees of freedom parameter imply substantial tail dependence. The tail dependence would have been missed if Gaussian models had been used, since there is no tail dependence in a multivariate normal distribution, except in the case of perfect correlation. The tail dependence also seems apparent in the data, for example in Figure 1 where individual curves that are extreme at one tract location tend to be extreme at other locations.

Our findings are consistent with previous studies of diffusivity. These have shown that acute axonal injury at first decreases parallel diffusivity, but, as the tissue gets cleaned up, parallel diffusivity normalizes and then increases above normal. This is likely due to increased overall diffusivity and is probably not specific for a particular kind of tissue injury, e.g., not specific to MS. Another possibility is that the researchers have overestimated parallel diffusivity, because as the tract atrophies, one get contributions from adjacent cerebrospinal fluid (where diffusivity is high). However, the investigators have tried to minimize this effect, and it not considered to be a major problem.

The main contributions of this new method to the understanding of MS are (1) identification of the most severely affected cases; and (2) identification of the interesting parts of the tract.

6 Conclusion

This paper introduces a model for functional data that exhibit non-Gaussian characteristics such as skewness, heavy tails, and tail dependence. Our model includes a Karhunen-Loève expansion for a latent, but estimable, Gaussian process that induces dependencies. Our approach is based on copula methodology and combines elements of parametric and non-parametric modeling. However, robustness of our methodology to the choice of the parametric families of marginal and copula distributions and goodness-of-fit testing of these families remain open problems.

7 Acknowledgment

The authors would like to acknowledge Peter Calabresi for providing the data and the National Multiple Sclerosis Society and EMD Serono for providing funds for data collection.

References

[1] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12**, 171-178.

[2] Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *J. of the Royal Statistics Society, Series B* **65**, 367–389.

[3] Azzalini, A. (2010). R package 'sn': The skew-normal and skew-t distributions (version 0.4-15). URL http://azzalini.stat.unipd.it/SN.

[4] Basser, P., Mattiello, J., and LeBihan, D. (1994). Mr diffusion tensor spectroscopy and imaging. *Biophysical Journal* **66**, 259–267.

[5] Basser, P., Pajevic, S., Pierpaoli, C., and Duda, J. (2000). In vivo fiber tractography using dt-mri data. *Magnetic Resonance in Medicine* **44**, 625–632.

[6] Brumback, B. A. and Rice, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). *Journal of the American Statistical Association* **93**, 961–976.

[7] Carroll, R.J., Ruppert, D. and Welsh, A. H. (1998). Local Estimating Equations. *Journal of the American Statistical Association* **93**, 214-227.

[8] Cuvelier, E. and Noirhomme-Fraiture, M. (2006). A Probability Distribution Of Functional Random Variable With A Functional Data Analysis Application. ICDM Workshops, 247–252.

[9] Crainiceanu C.M., Staicu, A.M. and Di, C. (2009). Generalized Multilevel Functional Regression. *Journal of the American Statistical Association* **104**, 1550–61.

[10] Di, C., Crainiceanu, C. M., Caffo, B. S. and Punjabi, N. M. (2009). Multilevel functional principal component analysis. *Annals of Applied Statistics* **3**, 458-488.

[11] Embrechts, P., Mcneil, A. and Straumann, D. (2002). *Risk Management: Value at Risk and Beyond*, ed. M.A.H. Dempster, Cambridge University Press, Cambridge, 176–223.

[12] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis : Theory and Practice*. Springer Series in Statistics, Springer, New York

[13] Fernandez, C., and Steel, M.F.J. (1998). On Bayesian Modelling of Fat Tails and Skewness, *J. of the American Statistical Association* **93**, 359–371.

[14] Genton, M.G. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, edited volume. Boca Raton, FL: Chapman & Hall/CRC.

[15] Greven, S., Crainiceanu, M. C., Caffo, B. and Reich, D. (2010). Longitudinal Functional Principal Component Analysis, *Electronic Journal of Statistics*, to appear.

[16] Goldsmith, J., Feder. J., Crainiceanu, C.M., Caffo, B., and Reich, D. (2010) Penalized functional regression. Johns Hopkins University, Dept. of Biostatistics Working Papers. Working Paper 204. http://www.bepress.com/jhubiostat/paper204

[17] Handcock, M. S. and Stein, M. L. (1993). A Bayesian analysis of kriging. *Technometrics* **35**, 403–410.

[18] Li, Y. and Ruppert, D. (2008). On The Asymptotics Of Penalized Splines, *Biometrika* **95**, 415–436.

[19] Kneip, A. and Sarda, P. (2010) Sélection de modèle inculant des componantes principales. *42èmes Journées de Statistiques*. Marseille, France. [**needs to be completed**]

[20] McNeil, A., Frey, R., and Embrechts, P. (2005) *Quantitative Risk Management*, Princeton University Press, Princeton and Oxford.

[21] Morris J. S. and Carroll R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society B.* **68**, 179–199.

[22] Nelsen, R.B. (2006). *An Introduction to Copulas*. Springer Series in Statistics, Springer, New York

[23] Ozturk A., Smith S.A., Gordon-Lipkin E.M., Harrison D.M., Shiee N., Pham D.L., Caffo B.S., Calabresi P.A., Reich D.S. (2010) MRI of the corpus callosum in multiple sclerosis: association with disability. *Multiple Sclerosis* **16**, 166–177.

[24] Pace, L., Salvan, A., 1997. *Principles of Statistical Inference from a Neo-Fisherian Perspective*. World Scientific Press, Singapore.

[25] Ramsay, J. O. and Silverman, B.W. (2006). *Functional Data Analysis*, 2nd ed., New York: Springer.

[26] Ramsay, J. O., Hooker, G., and Graves, S. (2009) *Functional Data Analysis with R and MATLAB*, New York: Springer.

[27] Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics* **11**, 735–757

[28] Ruppert, D., Wand, M., and Carroll, R. (2003). *Semiparametric Regression.* Cambridge University Press, Cambridge.

[29] Sartori, N. (2006). Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *Journal of Statistical Planning and Inference* **136**, 4259–75.

[30] Sklar A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ Inst Stat Univ Paris* **8**, 229–231.

[31] Staicu A.-M., Crainiceanu C.M., Carroll R.J. (2009). Fast Methods for Spatially Correlated Multilevel Functional Data. *Biostatistics*, to appear.

[32] Staniswalis, J.G. and Lee, J.J. (1998). Nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association* **93**, 1403–1418.

[33] Yaglom, A.M. (1987) *Correlation Theory of Stationary and Related Random Functions I. Basic Results*. New York: Springer-Verlag.

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

[35] Zhang J., Jones M., DeBoy C.A., Reich D.S., Farrell J.A., Hoffman P.N., Griffin J.W., Sheikh K.A., Miller M.I., Mori S. and Calabresi P.A. (2009). Diffusion tensor magnetic resonance imaging of Wallerian degeneration in rat spinal cord after dorsal root axotomy. *Journal of Neuroscience* **29**, 3160–3171.

[36] Zhao, Y. and Joe, H. (2005). Composite likelihood estimation in multivariate data analysis, *The Canadian Journal of Statistics* **33**, 335–56

[37] Wood, S. (2006). *Generalized Additive Models: An Introduction with R.* Boca Raton, Chapman and Hall/CRC.