Modeling Functional Data with Spatially Heterogeneous Shape Characteristics

Ana-Maria Staicu,1,* Ciprian M. Crainiceanu,2 Daniel S. Reich,3 and David Ruppert4

1Department of Statistics, North Carolina State University, 2311 Stinson Drive, Campus Box 8203, Raleigh, North Carolina 27695-8203, U.S.A.
2Department of Biostatistics, Johns Hopkins University, 615 North Wolfe Street, E3636 Baltimore, Maryland 21205, U.S.A.
3Translational Neuroradiology Unit, Neuroimmunology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, 10 Center Drive MSC 1400, Building 10 Room 5C103, Bethesda, Maryland 20892, U.S.A.
4Andrew Schultz, Jr., Professor of Engineering, School of Operational Research and Information Engineering, 1170 Comstock Hall, Cornell University, New York 14853-3801, U.S.A.

Summary. 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 magnetic resonance imaging (MRI).

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.

In this article, 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 (Crainiceanu, Staicu, and Di, 2009; Di et al., 2009; 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; in comparison the nonparametric 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 nonparametrically as function of location.

The third novel feature here is the use of copulas to model spatial dependence. We use Gaussian or $t$-copulas, because, 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 (DTI) Study

The methodology described in this article is motivated by a state-of-the-art DTI study that analyzes and compares white
Biometrics

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 gray curves. Distance units are arbitrary.

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, Mattiello, and LeBihan, 1994; Basser et al., 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 coaxial 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 a few ways, such as mean and variability along the tract, but perhaps even more so with respect to skewness. The MS subjects’ profiles exhibit more 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, because 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 FDA—see, for example, the excellent monographs of Ramsay and Silverman (2005), Ferraty and Vieu (2006), and Ramsay, Hooker, and Graves (2009). A fundamental idea in FDA is to decompose the space of curves into principal directions of variation by a 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 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; see Section 6 and Figure S2 of the supplemental materials.

The article is organized as follows. Section 2 introduces functional processes with heterogeneous shape characteristics and describes the modeling methodology. Section 3 develops the estimation procedure and Section 4 discusses prediction.
Section 5 contains an extensive simulation study. Section 6 presents a comparison of our methodology with alternatives. Applications to the tractography data are presented in Section 7. A brief discussion is given in Section 8.

2. Functional Models with Spatially Varying Marginal Distributions

Let \( \{Y_i(t_y); t_y \in T\} \) 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 \( T \) and sampled at a grid of points \( \{t_{i1}, \ldots, t_{im_i}\} \subset T \); typically we take \( T = [0, 1] \). Furthermore, suppose that

\[
Y_i(t) = \mu(t) + \sigma(t)G^{-1}\{W_i(t); a(t)\},
\]

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 \( G(\cdot; a) \) is a cumulative distribution function in a parametric family of distributions with zero mean, unit variance, and shape parameter vector \( a \). For example, \( a \) 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, \( G^{-1}(\cdot; a) \) denotes the usual inverse cumulative distribution function (CDF). We assume that the shape parameter \( a(t) \) varies smoothly with \( t \). To simplify notation, we sometimes use \( G_t := G(\cdot; a(t)) \). Our main objectives are to estimate the population functions \( \mu(t), \sigma(t), \) and \( a(t) \) 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 \( p \)th quantile: \( Q_p(t) = \mu(t) + \sigma(t)G^{-1}_p(\cdot), \quad 0 < p < 1 \). 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 \( W_i \). Because \( W_i(t) \) is uniformly distributed for each \( t \), the joint distribution of \( \{W_i(t_{11}), \ldots, W_i(t_{m_i})\} \) is a copula, which we will model parametrically using Gaussian or t copulas.

An alternative to our methodology is to estimate the marginal distributions and the correlation structure separately. The marginal distributions could be estimated, for example, by quantile regression or by fitting univariate distribution at each time point by maximum likelihood or methods of moments. The correlation structure could be estimated by PCA. What is missing with this approach is a technique for combining the marginal distributions and the correlation structure into a probability model for the joint distribution of the data. Our copula methodology supplies this missing technique.

2.1 Population-Level Parameters

We model the mean function \( \mu(t) \), variance function \( \sigma^2(t) \), and the shape parameter functions \( a(t) \) 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 \( B(t) = \{B_1(t), \ldots, B_q(t)\}^2 = \{1, t, \ldots, t^{q-1}, (t - \kappa_1)^q, \ldots, (t - \kappa_K)^q\} \) be a \( q \)th degree truncated power series spline basis with dimension \( q = 1 + p + K \) having fixed knots \( \kappa_1, \ldots, \kappa_K \). 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 \( \mu, \sigma, \) and \( a \), because some of these functions may require more knots than others. To keep notation simple, in our exposition we will use the same spline basis for each of these three functions. We write \( \mu(t) = B(t)^T\beta_{\mu} \) and \( \log[\sigma^2(t)] = B(t)^T\beta_{\sigma} \), where the spline bases and coefficient vectors are \( q \)-dimensional. In some cases it is desirable to model, not \( a(t) \), but a suitable transformation of \( a(t) \), say \( h(\cdot; a(t)) \), where \( h(\cdot; a(t)) \) applies a one-to-one transformation to each coordinate of \( a(t) \). For example, if any component of \( a(t) \) is positive we could log transform that component. Our model for \( a(t) \) is \( a(t) = h^{-1}(B(t)^T\beta_{a}) \). Here \( \beta_{a} \) is a \( q \times \dim(\cdot; a(t)) \) matrix of parameters.

2.2 Functional Dependence and Copulas

A copula is a multivariate distribution function whose univariate marginal distributions are all uniform on \((0,1)\). If \( X = (X_1, \ldots, X_q) \) is a random vector and if \( F_i \) is the continuous marginal distribution of \( X_i \) then the distribution of \( \{F_1(X_1), \ldots, F_q(X_q)\} \) is called the copula of \( X \). 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 \( X \) has a Gaussian copula, this does not imply that \( X \) is Gaussian, only that \( X \) has the same copula as some Gaussian random vector. Stated differently, \( X \) 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 \( \lim_{p \to 0} P(Y > F_Y^{-1}(p)|X > F_X^{-1}(p)) \) 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, and Straumann, 2002). This means that under a Gaussian copula, the components behave in the extreme tails as if 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 7.

3. Estimation Methodology

We use penalized maximum likelihood to estimate the mean, standard deviation, and shape parameter functions, \( \mu(t), \sigma(t), \) and \( a(t) \). Each of these functions contains a reasonably large
number of parameters and the correlation matrix of the Gaussian or f-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 multistage 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 \( t_j = t_j \) 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 \( \mu_j \), \( \sigma_j \), and \( \alpha_j \) are obtained by maximizing the pointwise likelihood function

\[
\ell_j(\mu_j, \sigma_j, \alpha_j) = \sum_{i=1}^{n_j} \log \left[ \frac{g(Y_{ij} - \mu_j)}{\sigma_j} \right] - \log(\sigma_j),
\]

where \( g(x; \alpha_j) = \partial G(x; \alpha_j) / \partial x \) is the density function corresponding to the distribution \( G_t \). For example, when \( G_t \) is in the skew normal or skew t family of distributions, the functions \texttt{sn.mle} or \texttt{st.mle} of the \texttt{sn} package for R (Azzalini, 2011) can compute these estimates.

Maximum likelihood estimation of \( \alpha \) 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 \( \alpha \) (the skewness parameter) 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 \( \alpha \) lead to small changes in the density. It should be noted, however, that infinite values of \( \alpha \) 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 I” below.

When the curves are sparsely sampled, another estimation method is needed. Note that the estimates of \( \mu_j \), \( \sigma_j \), and \( \alpha_j \) 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, because they will be smoothed further in Step 2. Suppose that the \( j \)th subject’s curve is observed at \( t_j \), \( j = 1, \ldots, m \). For estimation at \( t_j \) using \( L \)th degree local polynomial estimation, define \( \mu(t) = \sum_{l=0}^{L} \mu_l (t - t_j)^l \), \( \sigma(t) = \sum_{l=0}^{L} \sigma_l (t - t_j)^l \), and \( \alpha(t) = \sum_{l=0}^{L} \alpha_l (t - t_j)^l \) for the \( q \)th component function of \( \alpha(t) \). Then estimating equation approach minimizes

\[
\sum_{i=1}^{n} \sum_{j=1}^{m_j} \left[ \log \left[ g(Y_{ij} - \mu(t_j)) / \sigma(t_j) \right] \right] - \log(\sigma(t_j)) \right] K_{v}(t_j - t_k)/b, \]

over all coefficients in \( \mu(t) \), \( \sigma(t) \), and \( \alpha(t) \). The estimates of \( \mu \), \( \sigma \), and \( \alpha \) at \( t_k \) are \( \mu_0, \sigma_0 \), and \( (\alpha_0, \ldots, \alpha_0) \), respectively, where \( Q \) is the dimension of \( \alpha \). Here \( K \) is a kernel function and \( b \) is a bandwidth.

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 \( \hat{\mu}_j, \hat{\sigma}_j, \) and \( \hat{\alpha}_j, j = 1, \ldots, m \). Step 2 uses \( \hat{\mu}_j \) and \( \hat{\sigma}_j \) and, in Alternative I, \( \hat{\alpha}_j \) 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 \( \beta_\alpha \), where \( \mu(t) = B(t)^T \beta_\alpha \), is estimated by minimizing the penalty criterion

\[
PL(\beta_\alpha) = \sum_{j=1}^{m} \left( \hat{\mu}_j - B(t)^T \beta_\alpha \right)^2 + \lambda \Omega_\alpha(\beta_\alpha),
\]

where \( \Omega_\alpha(\beta_\alpha) = \beta_\alpha^T D_\alpha \beta_\alpha \). 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 \( \tilde{Y}_{ij} = (Y_{ij} - \hat{\mu}(t_j)) / \hat{\sigma}(t_j) \) be the standardized observations, using the estimates of the mean \( \hat{\mu}(\cdot) \) and variance functions \( \hat{\sigma}(\cdot) \) obtained above. Define the penalized marginal likelihood function:

\[
PL_\alpha(\beta_\alpha) = -2 \sum_{i=1}^{n} \sum_{j=1}^{m_j} \ell(\beta_\alpha; \tilde{Y}_{ij}) + \lambda \Omega_\alpha(\beta_\alpha),
\]

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

For the mean and variance estimation, we selected the smoothing parameters \( \lambda_\mu \) and \( \lambda_\sigma \) using the restricted maximum likelihood estimation (REML) (Wood, 2006). For the shape parameter function estimation \( \lambda_\alpha \) we used using the corrected Akaue information criterion (AIC) criterion (Ruppert et al., 2003); other choices, such as cross validation or AIC can be used. The function \texttt{optim} of the \texttt{stats} package for R with the quasi-Newton algorithm can be used for optimization. Let \( \hat{\mu}(t) = B(t)^T \tilde{\beta}_\alpha \), \( \hat{\sigma}(t) = \exp(B(t)^T \tilde{\beta}_\alpha / 2) \), and
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 \( \alpha \), we model transformed parameters \( h(\alpha) \) instead. For example, if \( G(\cdot; \alpha) \), \( \alpha \leq \infty \), is the skew normal distribution, then an appropriate transformation is \( h() = \log[\Phi(-s)/(1 - \Phi(-s))] \), the normal CDF with standard deviation \( s \), which maps \( \infty \) 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 \( \beta_\alpha \) minimizes the penalized criterion \( P\ell_a(\beta_\alpha) = \sum_{i=1}^{m-1} h(\hat{\alpha}_i) - B_j^2 \beta_\alpha^2 + \lambda \beta_\alpha^2 D_\alpha \beta_\alpha \), with \( \lambda \) and \( D_\alpha \) 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 \( C(\cdot; \Omega) \) be a copula distribution and \( c(\cdot; \Omega) \), the corresponding copula density, where \( \Omega \) is the vector of copula parameters. It follows that the log likelihood of the full model can be written as:

\[
\ell(\beta_\mu, \beta_\sigma, \beta_\alpha, \Omega) = \sum_{i=1}^{N} \sum_{j=1}^{m} \ell_j(\beta_\mu, \beta_\sigma, \beta_\alpha, Y_{ij}) + \sum_{i=1}^{N} \log[c_1(F_{Y_{i1}}(Y_{i1})), \ldots, F_{Y_{im}}(Y_{im}); \Omega)],
\]
where \( \ell_j(\beta_\mu, \beta_\sigma, \beta_\alpha, Y_{ij}) \) is the penalized log-likelihood function of \( Y_{ij} \) and \( F_{Y_{i1}}(Y_{i1}) = G(Y_{i1} - B_j^2 \beta_\alpha^2)/\exp(B_j^2 \beta_\sigma^2/2) \) is the penalized log-likelihood function of \( Y_{ij} \) and \( F_{Y_{i1}}(Y_{i1}) = G(Y_{i1} - B_j^2 \beta_\alpha^2)/\exp(B_j^2 \beta_\sigma^2/2) \) is the penalized log-likelihood function of \( Y_{ij} \).

Here \( B_j \) is \( B(t_j) \) for all \( i \) and \( j \). To avoid undersmoothing of the population-level functions, which may be caused by large-dimension function bases, a penalized criterion will be used. The mean, variance and shape parameter functions parameters \( \beta_\mu, \beta_\sigma, \beta_\alpha \) are chosen to minimize the penalized pseudo log likelihood \( P\ell(\beta_\mu, \beta_\sigma, \beta_\alpha, \Omega) = -2N^{-1}\ell(\beta_\mu, \beta_\sigma, \beta_\alpha, \Omega) + \lambda_\mu \Omega_\mu(\beta_\mu) + \lambda_\sigma \Omega_\sigma(\beta_\sigma) + \lambda_\alpha \Omega_\alpha(\beta_\alpha) \). The \( \Omega \) is a consistent non parametric estimate of the copula parameter \( \Omega \) and \( \Omega_\mu, \Omega_\sigma, \Omega_\alpha \) are roughness penalties. Estimation of \( \Omega \), that does not depend on parameters \( \beta \), is discussed in Section 3.2.

Minimization of \( P\ell(\beta_\mu, \beta_\sigma, \beta_\alpha, \Omega) \) does not have a closed-form solution. To avoid the challenging numerical optimization problem, we use a three-step iterative algorithm. Denote by \( \beta_\mu^{(k)}, \beta_\sigma^{(k)}, \beta_\alpha^{(k)} \) the current estimates of the mean parameters, variance parameters, and shape parameters, respectively. More specifically, in the first step the mean parameters \( \beta_\mu \) are updated by \( \beta_\mu^{(k+1)} = \arg\min_{\beta_\mu} \{-2N^{-1}\ell(\beta_\mu, \beta_\sigma^{(k)}, \beta_\alpha^{(k)}, \Omega) + \lambda_\mu \Omega_\mu(\beta_\mu)\} \).

In the second step the variance parameters \( \beta_\sigma \) are updated by \( \beta_\sigma^{(k+1)} = \arg\min_{\beta_\sigma} \{-2N^{-1}\ell(\beta_\mu^{(k)}, \beta_\sigma, \beta_\alpha^{(k)}, \Omega) + \lambda_\sigma \Omega_\sigma(\beta_\sigma)\} \).

Finally, in the third step the shape parameters \( \beta_\alpha \) are updated by \( \beta_\alpha^{(k+1)} = \arg\min_{\beta_\alpha} \{-2N^{-1}\ell(\beta_\mu^{(k)}, \beta_\sigma^{(k)}, \beta_\alpha, \Omega) + \lambda_\alpha \Omega_\alpha(\beta_\alpha)\} \). At each step the corresponding smoothing parameter is chosen using the corrected AIC criterion (Ruppert et al., 2003). These three steps are then iterated until convergence; estimates under independence can be used as initial estimates of mean, variance and shape parameters.

Our experience from simulation studies is that, provided that the copula parameters are estimated accurately, Alternative 2 estimates of the population functions \( \mu(\cdot), \sigma(\cdot), \alpha(\cdot) \) are not very different from the estimates that use the working independence assumption with regards to the bias and variability. However the computational cost of Alternative 2 is considerably higher.

3.2 Calibrating Copulas

The approach for estimating the copula parameters depends on the type of copula used. A Gaussian copula requires only estimation of the correlation matrix of the underlying Gaussian distribution. A \( \epsilon \)-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

\[
\rho(t) = \Phi^{-1}(G([Y(t) - \mu(t)]/\sigma(t); \alpha(t))].
\]

This process has standard normal marginal distributions, so \( K(s, t) = \text{cov}(Z(s), Z(t)) \) is both the process covariance and correlation function, and thus it is the latent copula correlation function. One straightforward way to estimate \( K(\cdot; \cdot) \) 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, \( \hat{Z}(t) = \Phi^{-1}(G([Y(t) - \mu(t)]/\sigma(t); \alpha(t))) \). We model \( \hat{Z}(t) \) as the sum of two independent components: a finite basis expansion process, say \( \hat{Z}(t) \), and a white noise process with variance \( \sigma_w^2(t = s) \). By a finite basis expansion process, we mean a Karhunen–Loève expansion truncated at a finite, and generally small, number of eigenvectors (Di et al., 2009).

Let \( \hat{G}(s, t) \) be the method of moment estimator of \( \text{cov}(\hat{Z}(s), \hat{Z}(t)) \). Let \( \hat{G}(s, t) = \hat{G}(s, t) \) for \( s \neq t \); for \( s = t \), \( \hat{G}(t, t) \) is estimated using a bivariate thin-plate spline smoother applied to \( \hat{G}(t, s) \) 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 \( \hat{Z} \) is modeled by a finite basis expansion process, it makes sense to use a reduced rank approximation of the covariance of \( \hat{Z} \). We estimate \( \sigma_w^2(t) \) by \( \hat{\sigma_w^2} = \int G(t, t - \epsilon) \epsilon dt \); if this estimate is not positive, then it is replaced by a small positive number. Assume that the reduced rank approximation of \( \hat{G} \) is \( \hat{G} = \hat{G}L \hat{\Psi}^T \), where \( \hat{\Psi} \) is an \( m \times L \) matrix with orthogonal columns.
Here \( L \) is a finite truncation, determined by say the percentage of explained variance (Di et al., 2009). The correlation matrix of the copula density is the correlation matrix corresponding to the covariance matrix \( \hat{G} + \hat{\sigma}^2I_m \). Because \( \hat{\sigma}^2 > 0 \), 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 \( t_0 = t_j \) 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 \( \tilde{Z}(t_j) \) gives the same result. Using (5.32) of McNeil et al. (2005), let \( \hat{\rho}(t_j, t_j') \) be the sample Kendall’s tau between \( \{Y_i(t_j), \ldots, Y_s(t_j)\} \) and \( \{Y_i(t_j'), \ldots, Y_s(t_j')\} \) and define \( \hat{G}(t_j, t_j') = \sin[\pi \hat{\rho}(t_j, t_j')]/2 \). Then, the estimation of \( K \) proceeds as above.

In our extensive simulation study, the two methods of estimating the copula correlation matrix performed almost identically. When the data are sparse and \( t_0 = t_j \) 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 article.

Suppose now that a \( t \)-copula is assumed. Consider the transformed random functions \( Z^*(t) = T(t)^{-1}G[Y_i(t) - \mu(t)]/\sigma(t); a(t)] \), where \( T(t) \) is the \( t \)-CDF with \( \nu \) degrees of freedom; these curves are distributed according to a \( t \)-process with \( \nu \) degrees of freedom. One easy way to understand the \( t \)-process is through the following explicit construction. Let \( \{Z_i(t): t \in T\} \) 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 \( K(t, \cdot) \) is also the correlation function.

Write \( Z_i^*(t) = Z_i(t)/\chi^2_m(\nu) \), where \( \chi^2_m(\nu) \) is a chi-squared random variable with \( \nu \) degrees of freedom and independent of \( Z_i \). If \( \nu > 2 \), then the covariance function of \( Z^*_i \) is \( K^*_i(s, t) = \nu/(\nu - 2)K(s, t) \) 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, \( K(\cdot, \cdot) \) 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 correlation matrix \( K \) and degrees of freedom parameter \( \nu \). Estimation of \( K \) using Kendall’s tau is as for Gaussian copulas.

To estimate \( \nu \), we use a “pseudo-likelihood” method. Pseudo-likelihood means estimating certain parameters by maximizing the likelihood with all other parameters fixed at estimates (Gong and Samaniego, 1981). Denote by \( \hat{K} \) the estimate of the latent correlation matrix. Let \( \hat{F}_j \) 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 \( \hat{W}_j = \hat{F}_j \{Y_i(t_j)\} = R_j/(N + 1) \), where \( R_j \) 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 \( \nu \) is to fit the \( t \)-copula to the \( \hat{W}_j \) with \( \hat{K} \) fixed at the estimate \( \hat{K} \).

4. Predicting Response Trajectories

Our methodology allows prediction of the trajectory \( Y \) for a subject from irregular measurements by prediction of the latent process \( W \). Let \( \hat{W}_i(t) \) be the predicted latent process at a new time point \( t \) and corresponding to the \( i \)th subject; the predicted response at time point \( t \) is determined as

\[
\hat{Y}_i(t) = \hat{\mu}(t) + \hat{\sigma}(t)G^{-1}\{\hat{W}_i(t); \hat{a}(t)\}.
\]

The prediction of the trajectory of the latent process differs according to the copula model assumed, and is detailed next.

Gaussian copulas. Consider first the case where the dependence is modeled with a Gaussian copula. Using (4) it follows that prediction of \( \hat{W}_i \) is equivalent to prediction of \( Z_i(t) = \Phi^{-1}\{W_i(t)\} \), or in fact of the denoised \( Z_i \). This is now easy, because under our model assumptions, \( Z_i(t) \) can be expressed using Karhunen–Loève expansion as \( Z_i(t) = \sum_{\ell=1}^\nu \psi_i(\ell)\lambda^{1/2}_\ell + \epsilon_i(t) \), where \( \{\psi_i: \ell \geq 1\} \) forms an orthogonal basis in the space of squared integrable functions, \( \lambda_\ell \) are uncorrelated, have mean zero and variance equal to one. Also \( \epsilon_i(t) \) is assumed to have mean zero, variance \( \sigma^2_i \) for all \( t \) and is independent of the \( \xi_s \).

Estimates of the covariance function of \( Z_i \)’s are obtained using the method of moments along with smoothing; see Section 3.2. Furthermore, estimates of the functional principal components (FPC) \( \psi_1 \)’s, of the eigenvalues \( \lambda_\ell \) and the noise variance are obtained by using Mercer’s theorem of the covariance function of the denoised curves. It follows that prediction of the trajectory \( \{Z_i(t)\} \) reduces to prediction of the FPC scores, \( \xi_\ell \). For this step we use Gaussian assumptions, that \( \xi_\ell \) and \( \epsilon_i(t) \) are jointly Gaussian. The best linear prediction of the FPC scores for the \( i \)th subject, given the data from that individual is the usual conditional expectation, which is found to be

\[
\hat{\xi}_\ell = E[\xi_\ell | Z_i] = \lambda^{1/2}_\ell \psi_\ell^T \Sigma^{-1} Z_i.
\]

Here \( Z_i = \{Z_i(t_1), \ldots, Z_i(t_m)\} \) is the transformed data for the \( i \)th individual, \( \psi_\ell = \{\psi(t_1), \ldots, \psi(t_m)\} \) and \( \Sigma_{Z_i} \) is the \( m \times m \) matrix of the covariance of \( Z_i \), with the \( (j, j) \)th entry equal to \( (\Sigma_{Z_i})_{jj} = m \sum_{\ell=1}^\nu \lambda_\ell^2 \psi_i(t_j) \psi_i(t_j) + \sigma^2_i(t_j) \).

Then the prediction of the denoised transformed trajectory \( Z_i(t) \) for the \( i \)th subject is \( \hat{Z}_i(t) = \sum_{\ell=1}^\nu \hat{\psi}_i(\ell)\hat{\lambda}_\ell \psi_i(t_j) + \hat{\epsilon}_i(t) \), using the first \( L \) eigenfunctions, where \( \hat{\psi}_1 \)’s and \( \hat{\lambda}_\ell \)’s are the estimated FPC eigenfunctions and eigenvalues, respectively, and \( \hat{\xi}_\ell \)’s are the estimated FPC scores. Here \( L \) is a finite truncation, determined by say the percentage of explained variance (Di et al., 2009). It follows that the prediction of \( \hat{W}_i(t) \) is \( \hat{W}_i(t) = \hat{Z}_i(t) \).

t-copulas. We consider the case of \( t \)-copula model and assume that the degree of freedom parameter, \( \nu \), is larger than 2. Now we need to work with \( Z_i(t) \) which require transformation of the data using \( T(t) \) distribution, instead of \( Z_i(t) \) which uses the standard normal transformation. Notice that

\[
Z_i(t) = \sum_{\ell=1}^\nu \lambda^{1/2}_\ell \psi_i(\ell) \xi_\ell + \epsilon_i(t), \quad \xi_\ell = \xi_\ell / \sqrt{\chi^2_m(\nu)/\nu}, \quad \epsilon_i(t) = \epsilon_i(t)/\sqrt{\chi^2_m(\nu)}
\]

and \( \lambda_\ell \)’s and \( \psi_i \) defined as before. Although the Gaussian assumptions no longer hold in this case, there is empirical evidence (Yao et al., 2005) that the above formula of \( \xi_\ell \) still provides the best linear prediction of the FPC scores, \( \xi_\ell \), given the transformed information for the \( i \)th subject. There are a few modifications to be made:
the new variance of the noise is \( \nu/(\nu - 2)\sigma^2 \) and (2) the new variance of the scores is \( \nu/(\nu - 2) \). The prediction for the denoised transformed trajectory \( Z_i(t) \) for the \( i \)th subject is then \( \hat{Z}_i(t) = \sum_{j=1}^{L_i} \psi_j(t) \hat{\lambda}_j \xi_j \) thus the prediction of \( W_i(t) \) is \( \hat{W}_i(t) = T^{-1} \{ \hat{Z}_i(t) \} \), where \( \hat{\nu} \) is the estimated degrees of freedom of the \( t \)-copula.

Prediction can be done with standard functional PCA tools; Section 5.2 contains the prediction errors obtained using conditional expectation directly (Di et al., 2009) and they are similar to the ones obtained by our method. However, our method also estimates the pointwise quantile functions, which is essential for prediction intervals. The pointwise quantile functions estimated using FPCA are based on a normality assumption of the scores and can fail terribly, as discussed in Section 6 and illustrated in the supplementary material. Prediction methods currently do not exist for quantile regression because dependence is not modeled. However, it is possible to combine quantile regression for modeling the marginal distributions with our copula methodology for modeling dependence, as mentioned at the end of Section 6. For these reasons we view our modeling approach as a unifying approach for both prediction and quantile estimation.

5. **Simulation Studies**

We conducted an extensive simulation study to assess the performance of the estimation procedures described in Section 3. For each of the combinations of mean, standard deviation, shape parameter, and covariance function described below, 100 data sets, each consisting of \( N = 200 \) random trajectories, were generated from model (1). Each curve was sampled on a grid of equispaced time points \( \{t_j: j = 1, \ldots, 80\} \) in [0, 1], where \( i = 1, \ldots, N \). The simulated processes use \( W_i(t) = F_{X_i}(X_i(t)) \), where \( X_i(t) \) is a sample of mean-zero Gaussian curves and \( F_{X_i}(\cdot) \) is the CDF of \( X_i(t) \) for all \( 0 \leq t \leq 1 \).

We set \( X_i(t) = Z_i(t) + \epsilon_i(t) \), where the \( \epsilon_i(t) \) are independent \( \mathcal{N}(0, \sigma^2 = 0.10) \). Two main covariance structures of the underlying process \( Z_i \) are considered:

(I) **Finite basis expansion.** We assume \( Z_i \)'s are a sample of random curves with mean 0 and covariance function \( K_2(s, t) = \text{cov}\{Z_i(s), Z_i(t)\} \). The covariance function has the expansion \( K_2(s, t) = \sum_{j=1}^{L_i} \lambda_j \psi_j(s) \psi_j(t) \) in terms of eigenfunctions \( \psi_j(s) \) and eigenvalues \( \lambda_j \)'s. In the simulations, \( L = 3 \) and there were two cases:

(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) \( \psi_1(t) = \sqrt{2}(2t^2 - 1) \), \( \psi_2(t) = \sqrt{5}(6t^2 - 6t + 1) \) and \( \psi_3(t) = \sqrt{7}(20t^3 - 30t^2 + 12t - 1) \), where \( 0 \leq t \leq 1 \). We chose \( \lambda_i = (1/2)^{j-1} \) for \( j = 1, 2, 3 \).

(II) **Matérn covariance structure.** We let \( Z_i \) be Gaussian process with mean zero, variance equal to 2, and Matérn autocorrelation function

\[
\rho(\Delta; \phi, \kappa) = \frac{1}{2^{1-\alpha}} \Gamma(\alpha) \left( \frac{2^{\alpha/2} \Delta}{\phi} \right)^\kappa K_{\alpha} \left( \frac{2^{\alpha/2} \Delta}{\phi} \right),
\]

with (i) range \( \phi = 0.07 \) and order \( \kappa = 1 \); and (ii) with range \( \phi = 0.14 \) and order \( \kappa = 1 \). Here \( K_{\alpha} \) is the modified Bessel function of order \( \kappa \).

### Table 1

<table>
<thead>
<tr>
<th>Population level functions</th>
<th>( \sqrt{\text{ISBIAS}} ) (( \times 10^2 ))</th>
<th>( \sqrt{\text{IVAR}} ) (( \times 10^3 ))</th>
<th>( \text{IMSE} ) (( \times 10^2 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(ii) Mean</td>
<td>0.27</td>
<td>0.68</td>
<td>0.72</td>
</tr>
<tr>
<td>(ii)</td>
<td>0.27</td>
<td>0.70</td>
<td>0.75</td>
</tr>
<tr>
<td>(ii)</td>
<td>0.27</td>
<td>0.55</td>
<td>0.61</td>
</tr>
<tr>
<td>(ii)</td>
<td>0.27</td>
<td>0.63</td>
<td>0.69</td>
</tr>
<tr>
<td>(ii) Log variance</td>
<td>1.71</td>
<td>10.06</td>
<td>10.16</td>
</tr>
<tr>
<td>(ii)</td>
<td>1.45</td>
<td>10.45</td>
<td>10.50</td>
</tr>
<tr>
<td>(ii)</td>
<td>1.69</td>
<td>7.58</td>
<td>7.73</td>
</tr>
<tr>
<td>(ii)</td>
<td>1.68</td>
<td>9.51</td>
<td>9.61</td>
</tr>
<tr>
<td>(ii) Shape</td>
<td>67.35</td>
<td>111.36</td>
<td>129.67</td>
</tr>
<tr>
<td>(ii)</td>
<td>66.44</td>
<td>112.39</td>
<td>130.07</td>
</tr>
<tr>
<td>(ii)</td>
<td>61.30</td>
<td>57.72</td>
<td>84.00</td>
</tr>
<tr>
<td>(ii)</td>
<td>65.54</td>
<td>71.76</td>
<td>96.25</td>
</tr>
</tbody>
</table>

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

1. **Mean function:** (a) \( \mu(t) = 6 \); and (b) \( \mu(t) = -2.2t^3 - 1.2t + 0.7 \).
2. **Variance function:** (a) \( \sigma^2(t) = \exp(-5) \); and (b) \( \sigma^2(t) = \{2.2t^3 - 1.2t + 0.3\}/\exp(4) \).
3. **Shape parameter function:** (a) \( \alpha(t) = 0 \); (b) \( \alpha(t) = -21(t \leq 0.5) + 41(t > 0.5) \); (c) \( \alpha(t) = 5t^2 - 19t + 5 \); and (d) \( \alpha(t) = -10\sin(2\pi t) \), for \( t \in [0, 1] \).

5.1 **Model Estimation Performance**

For the estimation of the mean and variance functions, the smoothing parameters were selected using REML implemented in the \texttt{R} package \texttt{mgcv} (Wood, 2006). For the estimation of the shape parameter functions, we used cubic regression splines with five 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).

The supplementary material (Figure S1) includes an illustration of a data set generated from covariance structure (IIii), mean function (1b), variance function (2b), and shape parameter function (3d) when the latent process is specified as in (Ii) and (Iii) or in (IIi) and (IIii).
Biometrics

much, as will be seen soon. In Table 1, we report estimates of the square root of the integrated mean squared error (IMSE), the integrated squared bias (ISBIAS), and the integrated variance (IVAR). For example, for \( \hat{\mu}(t) \), these quantities are defined as: IMSE = \( \int_0^1 E[(\hat{\mu}(t) - \mu(t))^2] \, dt \), IVAR = \( \int_0^1 \text{Var}[\hat{\mu}(t)] \, dt \) and ISBIAS = \( \int_0^1 [E(\hat{\mu}(t)) - \mu(t)]^2 \, dt \). Here \( E(\hat{\mu}(t)) \) and \( \text{Var}(\hat{\mu}(t)) \) denote the sample mean and sample variance of \( \hat{\mu}(t) \), 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 (III), better estimates are obtained for all the population functions.

Figure 2 illustrates the performance of our methodology with respect to capturing the true dependence of the functional process. Figure 2 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 defining 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.

5.2 Prediction Performance

We investigate the mean prediction error, to assess the performance of our prediction approach. Consider first a Gaussian

![Figure 2](image-url)
Estimates of the squared root of IPE using the observed data (IPEo) and corresponding to two incompletely observed trajectories: one at the first 41 of the sampling design points (IPEnew1) and one at 41 random time points (IPEnew2). In brackets are displayed the counterparts using common functional PCA tools (Di et al., 2009). The scenarios use skew normal distribution with the mean (1b), the log variance (2b), the shape parameter (3d), and covariance structures (Ii)–(Iii). The correct copula model is used in the estimation.

<table>
<thead>
<tr>
<th>Copula</th>
<th>$\sqrt{IPE_o}$</th>
<th>$\sqrt{IPE_{new1}}$</th>
<th>$\sqrt{IPE_{new2}}$</th>
<th>$\sqrt{IPE_0}$</th>
<th>$\sqrt{IPE_{new1}}$</th>
<th>$\sqrt{IPE_{new2}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(IIi)</td>
<td>0.029 (0.029)</td>
<td>0.030 (0.037)</td>
<td>0.028 (0.029)</td>
<td>0.023 (0.023)</td>
<td>0.023 (0.028)</td>
<td>0.022 (0.023)</td>
</tr>
<tr>
<td>(III)</td>
<td>0.056 (0.055)</td>
<td>0.125 (0.122)</td>
<td>0.075 (0.077)</td>
<td>0.045 (0.044)</td>
<td>0.094 (0.096)</td>
<td>0.056 (0.059)</td>
</tr>
<tr>
<td>(II)</td>
<td>0.025 (0.025)</td>
<td>0.029 (0.036)</td>
<td>0.025 (0.027)</td>
<td>0.020 (0.020)</td>
<td>0.023 (0.025)</td>
<td>0.020 (0.020)</td>
</tr>
<tr>
<td>(Ii)</td>
<td>0.038 (0.038)</td>
<td>0.139 (0.125)</td>
<td>0.049 (0.052)</td>
<td>0.031 (0.030)</td>
<td>0.096 (0.099)</td>
<td>0.037 (0.040)</td>
</tr>
</tbody>
</table>

6. Comparison between Quantile Regression and a Karhunen–Loève Expansion

One main advantage of our model is that it both estimates the pointwise quantile functions and provides a Karhunen–Loève (KL) expansion to model dependencies. We are not aware of any competing methodology with this feature. Nonetheless, nonparametric quantile regression and a KL might be considered as alternatives to our methodology, so a comparison with them is in order. For fixed $p \in (0, 1)$, our estimated quantile function is $\hat{Q}_p(t) = \hat{\mu}(t) + \hat{\sigma}(t)C^{-1}\{p; \hat{\alpha}(t)\}$. Table 3 gives the estimates of the IMSE and ISBIAS of $\hat{Q}_p(t)$ for $p = 1\%$, 10\%, and 50\%, for one particular scenario. These results are compared to quantile regression estimates using constrained linear splines with an AIC chosen roughness penalty, implemented

Table 2

<table>
<thead>
<tr>
<th>Method</th>
<th>Quantile level</th>
<th>$\sqrt{TSBIAS}$ ($\times 10^2$)</th>
<th>$\sqrt{TVAR}$ ($\times 10^2$)</th>
<th>$\sqrt{TMSE}$ ($\times 10^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model-based</td>
<td>1%</td>
<td>0.46</td>
<td>1.58</td>
<td>1.64</td>
</tr>
<tr>
<td>COBS</td>
<td>1%</td>
<td>1.40</td>
<td>1.95</td>
<td>2.39</td>
</tr>
<tr>
<td>Model-based</td>
<td>2%</td>
<td>0.40</td>
<td>1.40</td>
<td>1.45</td>
</tr>
<tr>
<td>COBS</td>
<td>2%</td>
<td>1.12</td>
<td>1.45</td>
<td>1.83</td>
</tr>
<tr>
<td>Model-based</td>
<td>5%</td>
<td>0.33</td>
<td>1.17</td>
<td>1.21</td>
</tr>
<tr>
<td>COBS</td>
<td>5%</td>
<td>0.71</td>
<td>1.18</td>
<td>1.37</td>
</tr>
<tr>
<td>Model-based</td>
<td>10%</td>
<td>0.30</td>
<td>0.99</td>
<td>1.03</td>
</tr>
<tr>
<td>COBS</td>
<td>10%</td>
<td>0.41</td>
<td>1.04</td>
<td>1.11</td>
</tr>
<tr>
<td>Model-based</td>
<td>50%</td>
<td>0.31</td>
<td>0.65</td>
<td>0.72</td>
</tr>
<tr>
<td>COBS</td>
<td>50%</td>
<td>0.39</td>
<td>0.77</td>
<td>0.85</td>
</tr>
</tbody>
</table>
Figure 3. Estimated quantiles of the parallel diffusivity profiles along 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 (gray 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.

Our method outperforms cobs with respect to ISBIAS, IVAR, and IMSE, especially for more extreme quantile levels. Of course, the modeling of dependence by the KL expansion is an advantage of our methodology over quantile regression, because the KL expansion provide a probability model for the data generating mechanism, which facilitates prediction (Section 4) and allows model-based bootstrapping. However, nonparametric regression quantile estimation is a good competitor with respect to marginal quantile estimation. For an additional comparison, in Figure S2 of the supplementary material, we plot for one scenario the quantile functions estimated with our method, the cobs software, and the KL expansion which assumes that the scores are normally distributed. As expected, the quantile functions with the latter method are badly biased when asymmetry exists in the data. Moreover, although the KL expansion can model correlation, because of its Gaussian assumption, the KL expansion cannot accommodate tail dependence.

Our methodology uses a parametric model for the marginal distributions, and we have focused on the skew normal and skew t families, with only three, respectively, four parameters. In situations where the marginal distributions are not well approximated by either of these models, one could use a more highly parameterized model, such as normal mixture models. Alternatively, one could use quantile regression regression to estimate the marginal distributions in combination with our copula approach to modeling dependence.

7. Application: DTI Tractography
We apply our proposed method to a study of the diffusion characteristics of white matter tracts in patients with MS and healthy controls. White matter tracts consist of axons that convey information between parts of the brain. These axons...
are covered with myelin, a white fatty coating. 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 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 uses measurements of the parallel diffusivity within the corpus callosum for 162 MS patients and 42 healthy controls. Goldsmith et al. (2011) 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 ($\nu$) was estimated through maximum likelihood under a working independence assumption. This approach yields $\hat{\nu} = 6.79$ for the MS group and $\hat{\nu} = 16.482.29$ for the control group. Because $\hat{\nu}$ of the control group is very large, and because the skew $t$ distribution reduces to skew normal when $\nu = \infty$, we used the skew normal for this group.

The estimated mean, variance, and skewness functions in the healthy volunteer and MS groups, along with confidence intervals of the corresponding difference functions, are included in the supplementary material, Figure S3. The results indicate slightly larger mean of parallel diffusivity for cases than for controls. Furthermore, using 90% bootstrap confidence intervals (based on 1000 resamples) they depict significantly higher variability and considerably larger third-moment skewness coefficient functions for the parallel diffusivity in the MS group than in the control group. 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. Figure S3 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 3 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 3 (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 4 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 these groups of curves, models with common correlation functions can be used in future analyses. At the same time, it indicates that the difference in the distributions of the parallel diffusivity profiles in the MS group and controls can be summarized solely by the marginal distributions. In particular one could use the pointwise quantile functions obtained in the control group to identify most unusual parallel diffusivity in the MS group. Furthermore, because the mean and the variance are significantly larger in the MS group than in the control group irrespective of the tract location, one can use the pointwise skewness to identify the most interesting parts of the tract.

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 to the Gaussian copula correlation matrices. The maximum likelihood estimates of copula degrees of freedom were 3.30 for the control group and 4.88 for the MS group. (These estimates should not be confused with the estimates of $\nu$ for the marginal skew t distributions.) Such small values of the copula degrees of freedom imply substantial tail dependence. The tail dependence would have been missed if Gaussian models had been used, because 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. This behavior would not occur if the data had a multivariate Gaussian distribution, even with high correlation, unless the correlation were perfect.

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 repairs, 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 gets 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.

8. Conclusion
This article 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.

9. Supplementary Material
Figures S1–S3 referenced in Sections 5–6 are available under the Paper Information link at the Biometrics website http://www.biometrics.tibs.org/.

Acknowledgements
We are grateful to the editor, associate editor, and one referee for their insightful comments. A-MS’s research was partially supported by NSF grant DMS 1007466. The research of CMC and DR was supported by NIH grant R01NS060910. The authors acknowledge Peter Calabresi for providing the data and the National Multiple Sclerosis Society and EMD Serono for providing funds for data collection.

References


Received November 2010. Revised July 2011. Accepted July 2011.